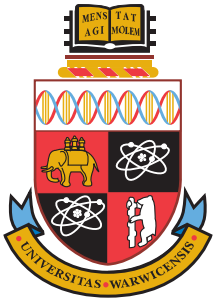

Understanding the "Rules of Engagement" for Membrane Protein Folding: Chemical Biology and Computational Approaches for Determination of Structure and Dynamics

by

Anthony Nash

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Supervisors: Dr. Ann Dixon and Dr. Rebecca Notman

MOAC Doctoral Training Centre

February 2014

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgements

I would like to begin by thanking the academics and staff at the MOAC doctoral training centre, without whom this PhD would never have been possible, nor nearly as enjoyable. My gratitude goes to the Centre for Scientific Computing for their excellent support and high performance computing facilities, the Chemical Biology department, and EPSRC for funding.

For three years I have been part of two wonderful groups of which I would like to thank: Leo Bowsher; Sang Young Noh; Jenny Webb; Michael Chow; Dr. Fay Probert, and Dr. Gemma Warren. Not only did they contribute towards a friendly and intellectually stimulating environment, but also many hours of fond and happy memories.

Both of my supervisors, Rebecca Notman and Ann Dixon, have contributed so much of their time and support above and beyond anything I have experienced during my previous jobs in industry. I cannot thank them enough for their guidance throughout my studies. In particular, I would like to thank Ann for introducing me to the world of Chemical Biology and experimental science. The knowledge and experience from the long hours spent in the laboratory has contributed vastly to my first steps towards my academic career. My other supervisor, Rebecca, I would like to thank for her time and energy and providing her support and knowledge without which I would not be in an academic career.

Finally, I would like to thank my wife, Dr. Johanna Nash. Johanna has been by my side throughout these three years, providing love, support, and what can only be described as an infinite amount of patience.

PhD published papers

- **Effects of the oncogenic $V_{664}E$ mutation on membrane insertion, structure, and sequence-dependent interactions of the neu transmembrane domain in Micelles and Model membranes: an integrated biophysical and simulation study.** *Andrew Beevers, Anthony Nash, Martha Salazar-Cancino, David Scott, Rebecca Notman and Ann Dixon.* Biochemistry. 2012. 51(12) 2558-2568.

Abstract

Approximately one third of genes in the human genome (1) encode transmembrane (TM) proteins and form more than half of all drug targets (2). However, our understanding of how these proteins fold into their functional form, as well as how they may misfold into a disease-associated form, remains a difficult area of study. By observing the effects of single point mutations in the context of a native sequence, in addition to adding and mutating interhelical interaction motifs on a low complexity sequence background, we aim to elicit 'rules' of TM protein domain association.

For the single point mutation in the context of a native sequence, the TM domain of the sequence Neu, along with its oncogenic substitution $V_{664}E$ form Neu*, were selected. Using molecular dynamics (MD) a united atom model of each dimer in a model bilayer system was subjected to umbrella sampling along an interhelical reaction coordinate to yield a free energy profile of self-association. The lipid order, bilayer thickness, and peptide tilt angle were calculated from trajectories taken from three points along the reaction coordinate. Helical composition, solvent accessible surface area, and hydrogen bond analysis (for the $V_{664}E$ substitution) were performed at the free energy minimum.

Low complexity sequences of polyleucine and polyleucine-alanine heptad repeat sequences, with and without interaction motifs similar to those present in the Neu model, were ligated into PBLM100 plasmids. Transformed *E. coli* cells were subjected to semi-quantitative homo-interaction analysis using the GALLEX assay. The same TM sequences were modelled using a coarse grained (CG) forcefield. Umbrella sampling along an interhelical reaction coordinate was performed to yield a free energy profile of self-association. Single-linkage cluster analysis of peptides was performed at the global free energy minimum. A representative structure from each set was compared to an averaged structure from the clusters of an atomistic conformational search.

The results presented in this study, could contribute to what in theory would be a large database of motif-driven rules for TM helical domain oligomerisation. This may encourage further investigation into TM protein design for novel application.

Chapter 1

Introduction

1.1 Statement of problem

TM protein folding, a fundamental process of cellular function and intercellular communication, remains a persistent nontrivial problem with the number of resolved structures of soluble proteins far out weighing those of membrane proteins (3). Predicting the protein structure from the primary sequence is fraught with difficulty. Our understanding of the process behind the spontaneous insertion of a TM protein and its folding is limited, with most information coming from the studies of native TM protein sequences in model membrane-mimetics. Unraveling the complex relationship between TM protein sequence and TM protein structure using low complexity sequences has been described as a 'divide and conquer' approach (4). By studying sequences of low complexity, putting to one side the often complex sequence of native proteins, we are able to improve our understanding of the interactions between α-helical domains. A low complexity sequence is composed from the smallest variation in amino acid types whilst maintaining a stable TM α-helical structure. The sequence can be used as a baseline for measuring TM helix-helix interaction and oligomerisation properties before and after changes to the primary sequence. It is often seen as a 'bottom-up' approach to understanding TM helix-helix interactions. Conversely, studying how a membrane protein folds from the perspective of a complete native TM domain sequence can be considered as 'top-down'. A low complexity sequence scaffold is the foundation for *de novo* TM α-helical peptide designs. Using the current understanding of how TM interactions drive association between peptides, the low complexity scaffold is modified to exhibit expected behaviour, for example,

designs for *de novo* membrane ion channels (5).

Our understanding of how TM proteins fold into their functional form, as well as how they may misfold into a disease-associated form, remains a difficult area of study. This is due primarily to the hydrophobic nature of TM proteins, which requires the use of membrane mimetics (e.g. detergent micelles and phospholipid vesicles) for their purification and further study in either the solution- or crystalline state. Many lipid mimetics may disrupt membrane protein structure and destabilise the protein (6), yielding poor quality data that is difficult to interpret. Likewise, studying *in silico* membrane model systems incorporating integral TM proteins can be fraught with long auto correlation times resulting in lengthy simulations.

Despite over three decades of research, the ability to predict membrane protein structure from the primary sequence is unreliable. Although it is thought that the primary sequence drives protein folding, the fold is subjected to the constraints imposed on the protein by the membrane bilayer architecture (7). Despite the experimental challenges, there have been many studies which suggest that membrane protein folding is dependent upon both protein-protein and protein-lipid interactions; therefore a greater understanding of both types of molecular interactions is crucial. By way of predicting the final structure of a membrane protein and in particular, of disease-associated folds, future therapeutic designs maybe within sight. Building *de novo* membrane proteins for a novel application is an emerging field, which will perhaps require novel approaches and new methodologies. As will be discussed in section 1.4, there are a far greater number of purpose built *de novo* soluble proteins than those of a membrane protein design.

We have investigated both a 'bottom-up' and 'top-down' approach using a combination of well established tools. Our investigation into TM protein folding from a 'bottom-up' perspective utilises *in vivo* experiments in cell membranes to investigate the propensity for oligomerisation of low complexity sequence scaffolds combined with *de novo* principles of helix-helix interactions to modulate the oligomerisation. The results are complemented by calculating the propensity for dimerisation of the low complexity sequences in model membranes using MD. The low 'bottom-up' study of low complexity sequences is accompanied by a 'top-down' approach using MD to investigate the effect of a single point mutation on a native TM domain sequence in a model membrane. PMF calculations are used to yield the free energy of association along a together-apart reaction coordinate between two *inter*helical faces. The mechanism of dimerisation is assessed by

studying the *inter*helical side chain contacts, the helix-helix crossing angle (see Appendix A for a formal definition), and analysing macroscopic changes to the bilayer with respect to the *inter*helical distance.

The following review begins with a brief discussion on the importance of TM proteins along with a short description of their secondary structure (section 1.2). This is complemented with a section on the characteristics of membrane architecture used in this study, and the studies discussed in this review (section 1.2.1). We then introduce an in depth review on helix-helix TM interactions motifs (section 1.2.2), followed by how oligomerisation of TM proteins are affected by the membrane bilayer architecture (section 1.2.3). This is then followed by a review on the limited number of low complexity TM protein designs in literature (section 1.3). The number of studies on low complexity sequences reflects the novel value of this study. Finally, the review is concluded with a discussion on the applications of *de novo* TM protein designs (section 1.4).

1.2 Transmembrane proteins

Approximately one third of genes in the human genome (1) encode TM proteins and form more than half of all drug targets (2). These proteins are able to associate with the cell membrane by either associating with the surface (monotopic) (Figure 1.1 (A)), or crossing the whole bilayer (integral) (Figure 1.1 (B)). When an integral TM protein spans the bilayer several times it is known as a polytopic integral TM protein (Figure 1.1 (C)). To associate with a bilayer, a monotopic TM protein typically requires a linear arrangement of hydrophobic residues to anchor amongst the hydrophobic lipid chains in the membrane core (due to the like-for-like arrangement of non-polar side chains with the non-polar alkyl structure of the lipid chains). Alternatively, the membrane protein can be anchored to the membrane by electrostatic interactions with the lipid head groups. In the case of an integral TM protein, the non-polar TM domain spans the non-polar bilayer core, with polar domains on either side of the membrane. The hydrophobic nature of the TM domain sequence prevents the protein from disassociating from the bilayer. In the case when a TM domain disassociates from the bilayer due to a disease-associated cleavage of the TM protein anchor or domain, in a multi-cellular organism globular aggregates can form in the extra-cellular matrix, potentially disrupting regular cellular function (8).

TM proteins that span the bilayer are typically involved in cell-signalling and assisted diffusion of

ions and small molecules. Extracellular membrane associating proteins on the other hand play a crucial role in direct cell-to-cell biological processes, such as the α-helical SNAP-25, part of the plasma membrane bound SNARE (Soluble *N*-ethylmaleimide sensitive fusion Attachment Protein Receptor) TM proteins, which drives exocytosis of transport vesicles in neuromediator release (9).

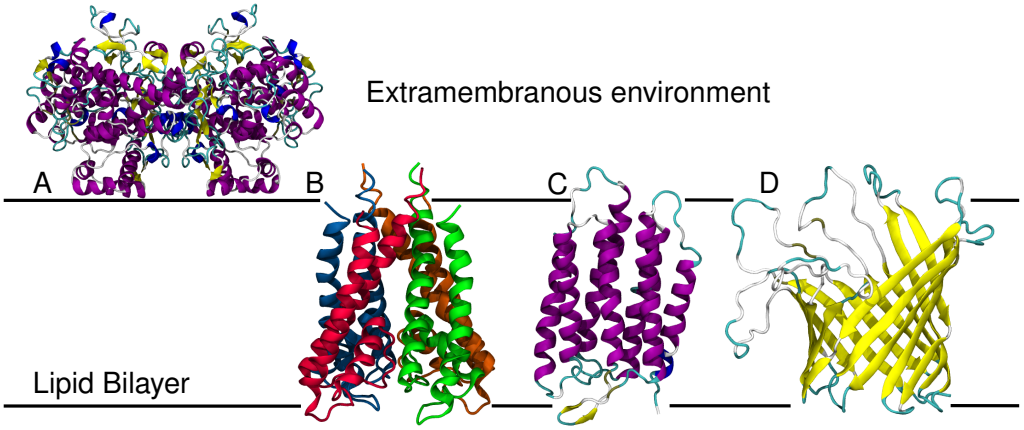

Figure 1.1: Structural types of membrane proteins: (A) Associating membrane protein: prostaglandin H2 synthase (associates to the membrane of the endoplasmic reticulum), 1PTH. (B) Integral TM protein: potassium channel KCSA, 1BL8. (C) Polytopic TM protein: Bacteriorhodopsin, 2AT9. (D) β-barrel Omp G monomeric porin 2JQY.

The α-helix is the most common secondary structure amongst TM protein domains. By adopting a right-handed spiral, residues are stacked up to four residues apart, which enable the formation of a network of hydrogen bonds between the amide hydrogen donor of residue i with the carbonyl oxygen acceptor of residue $i + 4$. This shields the polar peptide backbone from the non-polar lipid chains of the membrane. Helices are on average 26.4 nm in length and are tilted with respects to the bilayer normal by an average 21° (10). The tilt angle is defined as the angle between the principal axis of the helix with the bilayer normal. The orientation of the bilayer normal is known to fluctuate in a liquid crystalline state, however, the bilayer normal can be considered constant for the system in this study. For a formal definition of a helix tilt angle in a bilayer see Appendix A.

Although far less prevalent than α-helices, it is thought that 2-5% of the prokaryotic gram-negative genome encodes for β-barrel porin TM proteins (11) (Figure 1.1 (D)), and many TM proteins in mitochondrial and chloroplast outer membranes of eukaryotic cells are thought to adopt a β-barrel structure (12). The tertiary structure is constructed from an arrangement of anti-parallel β-strands, each consisting of three or more amino acids. Strands form a closed structure through an arrangement of hydrogen bonds between the peptide backbone amide hydrogen donor and carbonyl oxygen acceptor. TM β-barrel proteins are arranged in such a manner that the outside of

the structure is lined with non-polar side chains, whilst the inside is composed of polar side chains.

Two decades ago, a simple two-stage model of helical membrane protein spontaneous insertion and association was proposed (13). The first stage describes the spontaneous insertion of a hydrophobic protein into the bilayer as an independently stable α-helix, while the second stage describes the lateral association of helices within the plane of the bilayer (Figure 1.2). Although overly simplistic, the two-stage model is a reasonable attempt at describing the behaviour of several integral membrane proteins. A decade later, the two-stage model was extended by an additional step designed to encapsulate the events of ligand binding, folding of extramembranous loops, insertion of peripheral domains and the formation of quaternary structures (14). Over the last decade, work by Jean-Luc Popot has contributed towards designing chemical chaperons for the inserting and folding of membrane proteins. Hemifluorinated surfactants, constructed from a polar head and a hydrophobic moiety with a fluorinated central region with a hydrocarbon tip, have been shown to form a good solvent for membrane proteins whilst not mixing well with alcyl lipid chains (15). These surfactants promote further thermodynamic quantification of the mechanisms involved in chaperoning and spontaneously inserting membrane proteins.

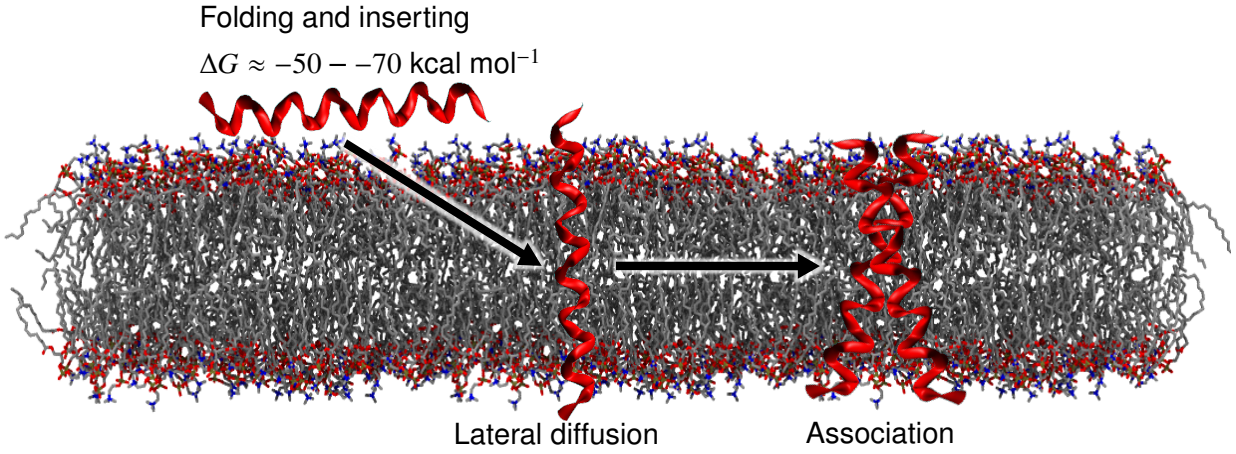

Figure 1.2: A pictorial representation of the membrane protein two-stage model. During the first stage, a hydrophobic protein spontaneously inserts into the bilayer as an independently stable α-helix. The initial model proposed that a 20 residue long polyalanine sequence is in favour of inserting by $\Delta G \approx -50 - -70$ kcal mol^{-1} (13). The second stage describes the lateral diffusion to eventual association with a second helix. Reversing the process of association and the evacuation of the protein from the membrane is usually energetically unfavourable and requires significant disruption to the bilayer, or release through proteolysis (8).

1.2.1 Membrane-mimetics

The following section briefly reviews the characteristics, uses, advantages, and disadvantages of membrane-mimetic systems discussed in this introduction and used throughout this study. There are already a significant number of in depth reviews into this topic (16, 17, 18); therefore, the aim of this subject is simply to introduce the reader to the terms used. The aforementioned points of interests are presented in Table 1.1.

1.2.1.1 Isotropic solvents

Some popular isotropic solvents are chloroform, trifluoroethanol (TFE), dimethyl sulfoxide (DMSO), methanol, ethanol, 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), and other hydrogenated or halogenated alcohols. They are the simplest and crudest hydrophobic membrane mimetic with a very poor resemblance to biological membranes due to the lack of chemical and structural heterogeneity, in addition to an absence in anisotropic characteristics. Unlike the more sophisticated membrane mimetics, such as lipid bilayer vesicles, isotropic solvents have a high solubility for TM peptides and they are applicable over a broad range of experimental techniques (as listed in Table 1.1).

1.2.1.2 Detergent micelles

With head group regions typically bulkier than their hydrocarbon chains, self-association into micelle structures shields the hydrophobic chains from the solvent environment by orientating detergent chains together whilst forming a spherical macroscopic structure. As a membrane mimetic, they are easy to prepare and use. With a moderate light scattering and fast tumbling, detergent micelles are ideal for CD and NMR, respectively. By the standard of today's high performance computers, micelles are small enough to be studied along with solvent and protein using an all-atom model (19).

Detergent micelles remain highly soluble for membrane proteins reconstitution. They are also known to adopt a cross-sectional topology similar to a lipid bilayer. Unfortunately, poor detergent packing, significant exchange between lipid and solvent, and disruption to protein tertiary structure, result in principal differences between detergent micelles and bilayers. The high curvature of detergent micelles is known to induce stress in a helical peptide compared to model bilayer

structures and natural bilayers (16).

1.2.1.3 Lipid bilayers

Model bilayer systems, whether as a bilayer slab or as part of a lipid vesicle (often referred to as a liposome) introduce the anisotropy and structural characteristics similar to those of natural bilayers. Early planar membrane model systems were composed of a monolayer at the air-water interface (20). By evaporating organic solvent, lipids are left to orientate their polar heads into the water leaving the hydrophilic chains orientated into the air. The precise surface pressure, thickness and phase can be controlled, although high surface tension of water can denature proteins.

The more familiar unilamellar lipid bilayers are constructed from two opposite facing lipid leaflets, whilst multilamellar lipid bilayers are constructed from stacked bilayer sheets, separated by a polar solvent. Although the curvature of a vesicle is far less pronounced than a detergent micelle, more lipids pack in the outer membrane of a vesicle to equalise surface tension between the two leaflets (21). Mixed lipid bilayers can represent a closer abstraction to natural bilayers and enables the calculation of changes to lipid aggregation and TM peptide packing by varying the head group charge, the bilayer thickness, and lateral pressure profiles (mediated through a change in area per lipid). Despite their obvious advantages and the relative ease in their construction, reconstituting membrane proteins into phospholipid vesicles requires a lengthy process of dialysing or gel filtering the detergent so as to solubilise the membrane proteins. In addition, there is little control over the orientation of the membrane protein across the bilayer (22). Due to the size of the vesicle (the diameter is typically > 50 nm), the system is too large as an all-atom model for even the most sophisticated computing platform.

1.2.1.4 Biological membranes

Biological membranes are complex and often exhibit unpredictable behaviour. They are crowded with membrane proteins, glycoproteins, and steroids, all of which can interfere with membrane protein function and structure by contributing to a life-time of non-specific interactions, thus making free energy calculation of association between membrane proteins of interest an approximation at best. The variation in lipid composition can also contribute to a change in membrane protein behaviour and structure. Cell-based TM protein interaction assays used for screening helix-helix

interactions, such as ToxR (23, 24), TOXCAT (25), GALLEX (26), and POSSYYCAT (27), are not always reproducible and are prone to experimental fluctuations. Further more, it is possible for a very small change in the environment, for example: heat shock; starvation; and resistance to or absence of antibiotics, can to lead to unpredicted cellular responses.

		Membrane-mimetics	
Advantages	**Disadvantages**	**Applicable**	**Not applicable**
Isotropic solvents			
Very easy to use	No anisotropy	AUC, DC, cross-linking,	CR, cell-based assays,
High solubility of TM regions	Poor solubility of hydrophilic flanks	DSC, IR, ITC, MD, NMR,	enzymatic assays, EPR,
Cheap	Disruption of tertiary structure	SEC (FI, DLS)	Mi, PAGE, SPR
Detergent micelles			
Easy to prepare and use	Poor detergent packing	AUC, CD, crossing-linking,	CR, cell-based assays, Mi
High solubility	Pronounced dynamics	enzymatic assays, EPR,	
Moderate light scattering	Changes in micellar size and shape	FI, IR, MD, NMR, PAGE,	
Fast tumbling	Disturbance of tertiary structure	(DLS, SEC, SPR)	
Lipid bilayers			
Native-like environment	Reconstitution necessary	CD, crossing-linking,	Cell-based assays,
Bilayer asymmetry possible	Substantial light scattering	enzymatic assays,	PAGE, SEC
Transmembrane transport measurable	Slow tumbling	EPR, FI, IR, ITC, MD,	
	Expensive	NMR, CR, SPR (AUC, DLS, DSC, Mi)	
Biological membranes			
Native environment	High complexity	CR, cell-based assays,	AUC, CD, DLS, DSC,
Organismic level	Difficult to control	crossing-linking,	IR, ITC, EPR, MD,
Transmembrane transport measurable	Interference with other cell components/functions	enzymatic assays, FI,	PAGE, SEC, SPR
	No high-resolution structures	Mi (NMR)	

Table 1.1: Recreated from a review on membrane-mimetics by Bordag and Keller (17). CR, conductance recordings; DLS, dynamic light calorimetry; EPR, electron paramagnetic resonance; FI, fluorescence-based methods; Mi, microscopy; SPR, surface plasmon resonance; CD, circular dichroism; MD, molecular dynamics; IR, infra-red-based methods; AUC, analytical ultracentrifuge; ITC, isothermal titration calorimetry; NMR, nuclear magnetic resonance; PAGE, polyacrylimide gel electropharisis methods

1.2.2 TM interaction motifs

TM domain interaction motifs are patterns of amino acids, key to mediating the strong association of TM domains. Alteration to any part of the motif results in significant disruption to TM domain association. With an identifiable pattern, motifs have been used as determinants of protein folding to help predict TM domain interactions and structure. Motifs identified in TM proteins fall into one of three categories: small-xxx-small, notably the GxxxG (GG4) motif; heptad repeats; or interactions

driven by the promiscuous positioning of polar side chains in the lipid chain environment. We look at each motif in detail below.

1.2.2.1 Small-xxx-small TM interaction motifs

Association of TM helices within a membrane bilayer can be stabilized by interactions between specific amino acid motifs. Statistical analysis of amino acid patterns of a Swiss-Prot database of helical TM proteins revealed an over representation of small residues (alanine, glycine and serine) at *i* and *i+4* in association with larger aliphatic residues (28). An interaction motif of this design is commonly referred to as small-xxx-small. A survey of known membrane protein sequences has shown that the GG4 motif is over represented in the TM domains of membrane proteins (29, 30).

Based on mutagenesis of the glycophorin A (GpA) TM helix, Lemmon *et al.,* (31) proposed that the sequence ^{75}L^{76}Ixx^{79}G^{80}Vxx^{83}G^{84}Vxx^{87}T helped establish self-association (see Figure 1.3 (A)). The arrangement of two glycines three residues apart (^{79}Gxxx^{83}G) results in an indentation along the same helical face. A conformational search in a low dielectric environment (32), confirmed by NMR structural analysis (30), concluded that GpA adopted a right handed crossing angle of -40° with closest point of helix-crossing at the ^{79}G and ^{83}G residues. The tight packing between *inter*helical glycines maximise van der Waals interactions, and a high crossing angle ensures additional enthalpic contributions between bulkier β-branched amino acids. By systematically adjusting the number of intermediate residues, three intermediate residues between glycines were confirmed as the optimum (24). Further analysis determined that the strong helix-helix dimerisation could be disrupted after substituting a glycine for an isoleucine. The bulkier hydrophobic side chain disrupts the tight packing arrangement, dramatically destabilising TM domains interactions. This was confirmed on low complexity scaffolds of polyvaline and polymethionine TM proteins using the TOXR assay (24). Since these early studies, GpA and the G83I mutant have become common standards for a number of *in vivo* and *in vitro* TM protein association assays (26, 33).

Besides GpA, the GG4 motif has been identified in the TM domain of several other integral membrane proteins. To name a few: the phage M12 coat protein (34); the heparan sulfate proteoglycan N-syndecan (35); the β_2-adrenergic receptor (36); the amyloid precursor protein β-carboxy-terminal fragment (8); and Japanese encephalitis virus prM protein adopting a heterodimer with E proteins (37). Small-xxx-small motifs have also been found in homodimers of: Scavenger receptor

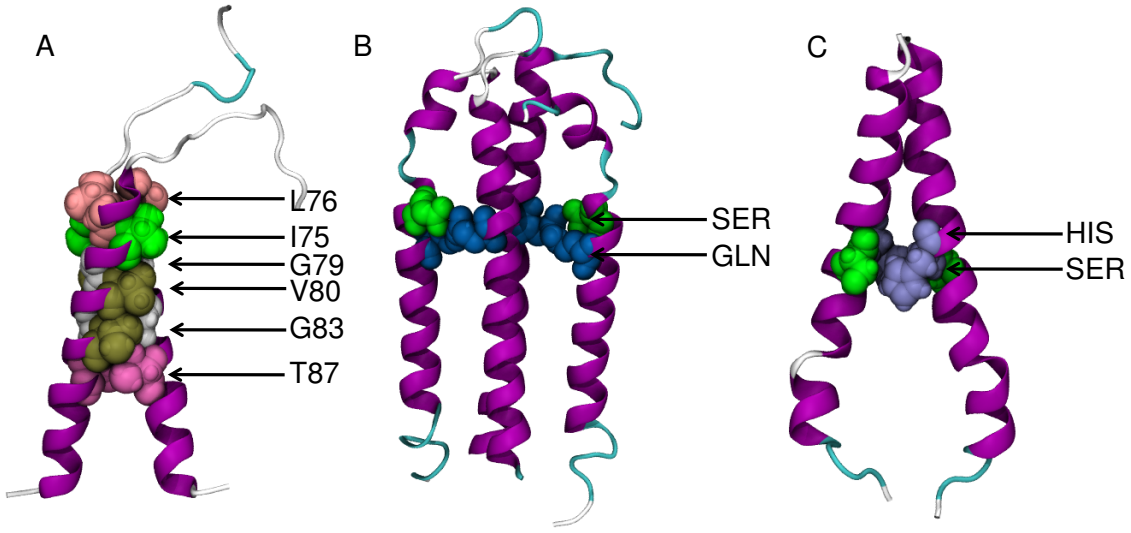

Figure 1.3: (A) The established minimum residues required to stabilise GpA TM domain. (B) Tar-1 (1VLT) TM domain showing polar residues serine and glutamine packing at the helix-helix interface. (C) BNIP3 (2KA1), small residues ^{176}A, ^{180}G and ^{184}G provide a close packing of helices which results in a hydrogen bond between serine and histidine.

class B, type I (38); ErbB2 (39); EphA1 (40); BNIP3 (41, 42) (Figure 1.3 (C)), and in the heterodimers of αIIb/β3 (43, 44) and ErbB1/ErbB3 (45). Although, neither isoleucine or valine are considered to have a short side chain, the motifs $I_{659}xxxV_{663}$ and $A_{661}xxxG_{665}$, have been found in rat Neu receptor tyrosine kinase TM domain (46).

Although the GG4 motif has been shown to play a significant role in TM helix dimerisation, there are claims that it is in fact the neighbouring aliphatic side chains, which stabilise TM domain association (47, 48), suggesting that the sequence modulates the strength of dimerisation. This was emphasised by Unterreitmeier *et al* (49), who demonstrated that dimerisation of a sequences containing a GG4 motif could be affected by a phenylalanine, yielding the sequence FxxGxxxG. An extension to the works of sequence context modulation of GG4 demonstrated histidine flanked with either glycine or serine stabilised GG4 mediated dimersiation (50). However, a minimised GpA motif (GVxxGVxxT) on a polyleucine peptide in detergent micelles yielded oligomerisation without the assistance of neighbouring side chains (51), making it apparent that the membrane mimetic used to solubilise the TM peptide also has a profound influence on oligomerisation.

1.2.2.2 Heptad repeat transmembrane interaction motifs

The heptad repeat motif, composed of a seven residue repeat *abcdefg*, is prevalent amongst integral membrane proteins and soluble proteins alike (52, 53). Its left handed helix-helix crossing angle

ensures contact between helices every 3.5 amino acids resulting in the seven amino acid repeat. On rare exceptions, a heptad repeat containing a GG4 motif can form a right handed helix-helix crossing angle (54), which is often a characteristic of a small-xxx-small mediated TM domain dimerisation.

The tight packing due to reduced residue turns in a left-handed helix-helix dimer results in a smaller positive crossing angle, giving the appearance of a 'coiled coil' or 'supercoiling'. The heptad pattern repeats two to three times, with a 'knobs-into-holes' arrangement of side chains stabilising oligomerisation (54). The side chains of one helix protrude into cavities formed by the side chains of the opposite helix typically using sites *a* and *d* as primary interaction sites (Figure 1.4 (A)). Side chains at sites *e* and *g* can be charged, resulting in *inter*helical salt-bridges. Finally, sites *b*, *c* and *f* are considered as sites interacting directly with the lipid environment. In the case of a TM trimer the same interaction sites contribute to oligomeric stability, however interactions are shared between three helical faces (Figure 1.4 (B)).

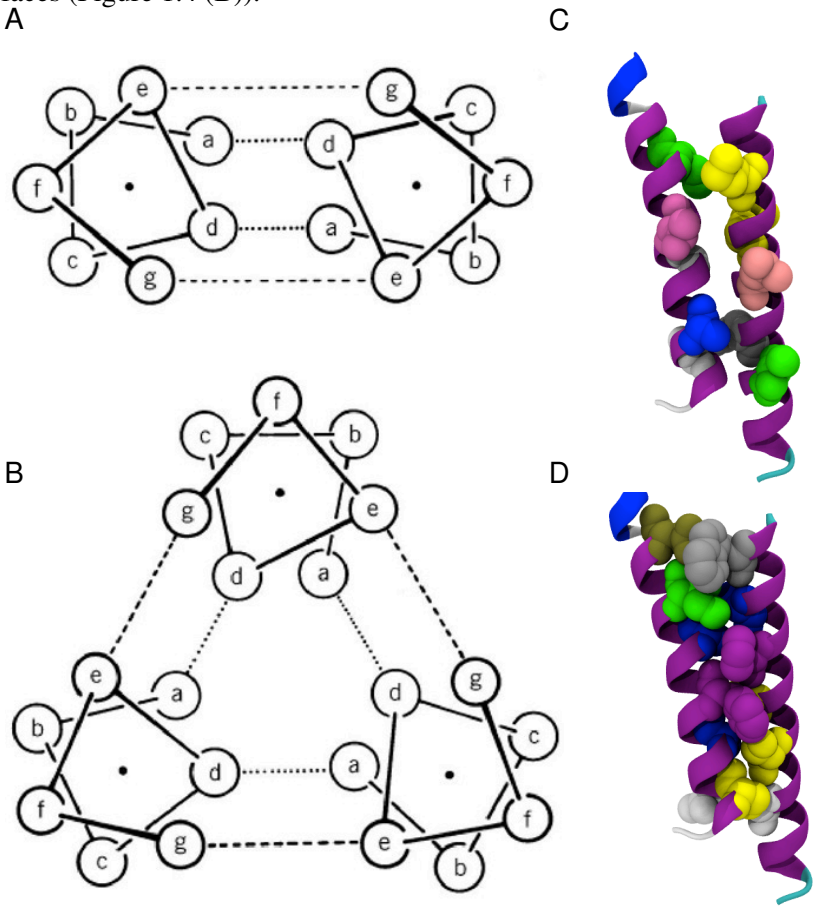

Figure 1.4: Heptad repeat motifs to support a TM (A) dimer, and (B) trimer. Recreated (55). The reaction centre from Blastochloris viridis strain (3T6D) (56), subunit M, showing TM domain heptad repeat (C) *g* and *e*, and (D) *a* and *d*, interaction sites.

Leucine zipper motifs	
Swiss-prot ID	**TM domain sequence**
CAD1_XENLA	^{705}I**L**GGI**L**A**LL**LL**LL**L**L**
CAD3_HUMAN	^{659}V**L**GAV**L**A**LL**FL**LL**V**LL**
CADB_CHICK	^{559}V**L**AV**L**GAV**L**AL**LL**V**LL**
CADF_HUMAN	^{611}**L**ASA**LLL**VLV**LL**VAL
CD72_MOUSE	^{93}**L**QNF**LL**G**LL**LSC**L**M**L**G
ENV_FRSFB	^{339}**LL**II**LLLLL**IL**LL**WT**L**
EPOR_MOUSE	^{256}**L**ILV**L**IS**LL**LTV**L**A**LL**
GPBB_HUMAN	^{155}**L**ALLG**L**G**LL**HA**LLL**V**L**
HEMA_CDVO	^{38}**LL**FV**LL**I**LL**VGI**L**A**LL**
LECH_CHICK	^{27}AVYV**LL**ALSFL**LL**T**LL**
PVR_MOUSE	^{35}**LL**VL**LL**AGGFLA**L**I**LL**
SRPB_MOUSE	^{35}**LL**SVAVA**LL**AV**LL**T**L**V
TNRC_MOUSE	^{223}**LL**AI**LL**S**L**VLF**LL**FTT
VE5A_BPV1	^{14}AAMQ**LLLLL**FL**LL**FF**L**
VGLX_HSVBS	^{395}**L**AIA**LL**V**LL**FS**L**VIV**L**

Table 1.2: Leucine zipper-like heptad repeats found in native TM domains. Reproduced from Gurezka (60).

Early studies in TM domain oligomerisation, found heptad repeat *inter*helical 'knobs-into-holes' packing in photosynthetic reaction centres (57) (Figure 1.4 (C) and (D)), bacteriorhodopsin (58), and cytochrome C oxidase (59). Further more, a recent statistical analysis on sequences from a Swiss-Prot database of TM proteins identified fifteen TM domains (see Table 1.2) which conformed to a sequence akin to a soluble leucine-zipper (60) (a soluble coiled-coil protein stabilised from the packing arrangement of stacked leucine side chains).

1.2.2.3 Polar transmembrane interaction motifs

The central region of the helical peptide adjacent to the lipid hydrocarbon region is occupied predominately by the nonpolar residues alanine, leucine, valine, isoleucine or phenylalanine. Polar or charged residues are usually found at the flanking region of the TM domain, the lipid-to-water interface, where residues such as arginine, aspartic acid, asparagine, glutamic acid, glutamine and lysine, have the potential to form O–H··O or N–H··O hydrogen bonds with lipid head groups regions of the bilayer or O-H$\cdots$O hydrogen bonding with water molecules (61). The behaviour described as the interaction between positively charged side chains and the negatively charged phosphate group of lipids is known as snorkelling (62). Water molecules are able to permeate the hydrophobic region of the bilayer by forming a network of hydrogen bonds from the charged side chain through to the bulk water. Besides interacting with water, charged side chains are thought to form salt-bridges,

which moderately contribute to helix-helix stability (7).

There is a significant energetic cost of inserting polar groups into the bilayer due to the unfavourable dipolar mismatch between nonpolar hydrocarbon chains and polar side chains (and polar backbones for a non-helical structure). Peptide and protein hydrophobicity scales (63, 64, 65, 66) estimate an energetic cost of 0.3 – 1.0 kcal/mol for the insertion of polar groups into the bilayer, and 1 – 6 kcal/mol for the insertion of peptide bonds or charged side chains. This penalty, however, has been attributed to some interesting mechanisms of TM protein association mediated through polar residue interactions.

One of the early works on polar residue mediated self association demonstrated the formation of a TM trimer upon the substitution of valine to asparagine on a low complexity sequence of a hydrophobic heptad repeat (67). Mutation of asparagine to valine essentially eliminates oligomerisation. This would suggest that a single asparagine residue buried at the centre of a very non-polar heptad repeat could provide the thermodynamic drive into membrane protein self-association. Interestingly, the heptad repeat sequence was derived from the soluble GCN4-P1 two stranded coiled-coil peptide, part of the yeast transcription factor GCN4 (68). This demonstrates the viability of retaining the buried hydrophobic residues of a soluble heptad repeat sequence for membrane protein designs. Gratkowski *et al.,* (69) extended this study by substituting the asparagine for glutamic acid, glutamine, lysine, serine and threonine. They concluded that glutamine, glutamic acid and histidine were capable of forming trimers due to a hydrogen bond network between polar side chains after protonation of side chain carboxyl groups. Serine and threonine were thought not to promote helix-helix association as the short side chain length would favour a hydrogen bond from the side chain hydroxyl group to the backbone carbonyl oxygen of the residue one helical turn directly above. Finally, it was concluded that lysine did not form helical bundles as it only has one polar atom (unlike the two polar groups on the side chain of glutamic acid and asparagine); therefore it had fewer sites to form *inter*helical hydrogen bonds.

A similar study by Zhou *et al.,* (70) demonstrated that polyleucine self-association could be driven by the inclusion of a singe glutamic acid, glutamine, histidine, aspartic acid and asparagine, yet threonine, and serine do not associate more than the polyleucine sequences. Interestingly, the serine motifs SxxSSxxT and SxxxSSxxT were later found to stabilise TM oligomerisation (71). It is likely that the packing of serine residues adopts a small-xxx-small alignment, bringing the helices closer

than in the case of a single serine residue, and thus encouraging *inter*helical hydrogen bonding. A further example of the role played by polar residues is the glutamic acid involved in TM protein oligomerisation of the oncogenic rat Neu and human ErbB-2, which has inspired a significant number of studies (72, 73, 74, 46), and still remains in contention to this day. A further review of this TM protein is given in chapter 4.

As it turns out, placing a polar residue into a TM sequence is not always enough to promote oligomerisation. The propensity for helix-helix association is complicated further by the specific location of the polar residue with respect to the hydrophobic core of the bilayer and how sequence context can modulate polar side chain *inter*helical association. Lear *et al.,* (75) performed the first study on the position of asparagine on a heptad repeat template in detergent micelles to elicit free energy differences using analytical ultracentrifugation. They concluded that an asparagine side chain within the hydrophobic core of the transmembrane dimer significantly stabilised self-association compared to when placed near the lipid-water interface. This would suggest the ability to shield the polarity of asparagine by packing beside the lipid-water interface (head group region). Further more, using *in vitro* translation of a model protein in the presence of dog pancreas rough microsomes, Hessa *et al.,* (66) determined such position-specific free energy contributions from all twenty amino acids. Strongly polar side chains were found to contribute approximately 1.8 kcal mol^{-1}, +1.8 kcal mol^{-1}, and +1.6 kcal mol^{-1}, per aspartic acid, lysine and asparagine residue, respectively. Statistical analysis over a data set of 170 integral membrane proteins (76), revealed that charged amino acids predominately favoured the lipid-water interface, a consequence of the energetic penalty of inserting charged amino acids into membrane bilayers. Further more, the distribution of charged amino acids was asymmetric, occurring more frequently on the cytoplasmic side of the TM helix. This asymmetric distribution of charged residues maybe due to the enrichment of sphingolipids and sterols in the extra-cytoplasmic leaflet (77). As far as the neighbouring sequence is concerned, Dawson *et al.,* (78), found that the effectiveness in self-association from *inter*helical hydrogen bonding between strongly polar residues was greatly influenced by the sequence of a native TM domain (notably, variants of bacteriophase M13 major coat protein, activated T cell receptor TM domains, and B cell antigen receptor domains), and Herrmann *et al.,* (79) concluded that glutamic acid, asparagine, lysine and arganine were significantly over represented in the presence of GG4 motifs. This would suggest that a single GG4 motif is usually not enough to stabilise TM domain

association alone, requiring the additional *inter*helical interactions of neighbouring polar residues, a further note of contention with regards to the precise dimerisation contribution a GG4 motif can make by itself.

1.2.3 Protein-lipid mediated oligomerisation

When considering the TM protein folding problem in a membrane environment, there are two things which should be considered: (i) what affect does the membrane-mimetic have on the intrinsic fold of individual membrane helices; and (ii) how does the membrane-mimetic composition affect helical TM domain aggregation. It is thought that as many as one in four TM domains lack the predicted hydrophobicity to allow spontaneous insertion into the bilayer (66). This suggests that the membrane protein folding problem is not simply a case of understanding TM helix-helix interactions, and that the lipid environment is thought to play a role in TM protein stability (80). Therefore, the choice in the membrane mimetic must be carefully considered whilst analysing the dynamic and structure of TM proteins. A recent study by Tulumello and Deber (80) suggested that a number of detergent micelles were capable of maintaining the helical composition of a TM segment similar to the helical structure of the intact protein in a native lipid environment. Interestingly, investigation of GpA in model membranes concluded that strong dimerisation of GpA during *in vitro* studies may be an artifact of homogeneous model bilayers, given that natural membrane environments promote destabilising contributions from electrostatic interactions between charged lipids and charged side chains and nonspecific competition from other membrane proteins (81). Generally speaking, a model bilayer is typically composed of one to three lipid types and has a large lipid to protein ratio. On the other hand, almost over half the total mass of natural membrane systems is contributed to membrane proteins (82). One would expect model membranes to promote a very different equilibrium for TM domain association.

It is well known that the chemical composition of the bilayer can lead to a condition of hydrophobic mismatch between TM helix length and bilayer thickness which, in turn, can induce conformational changes in a protein (83). The idea that the lipid environment may control the dimerisation mode of TM proteins is not new (84). Unrestrained united atom MD simulations and structural analysis using solid state NMR confirmed that a change from long chain lipids to short chain lipids increased the dimer tilt angle and resulted in a reduction in dimer stability (85). Such rearrangement can,

however, alter the TM domain secondary structure (86), and cases of hydrophobic mismatch have been shown to influence association of helical TM domains (87, 88, 89). Alternatively, there are many examples where the presence of protein has induced changes to the bilayer structure (see Figure 6.1 and Figure 6.2). For example, TM proteins can adjust to mismatch by stretching or disordering nearby lipids (90, 91, 83). Also, short helices can cause negative membrane curvature to the bilayer (92, 93, 94), and there is the suggestion that lipids are arranged according to the hydrophobic length of the TM domain (95). Polyansky *et al.,* demonstrated using computational free energy calculations, that the interactions between individual TM helices with different membrane-mimetics contributed to either an energy loss or energy gain which would consequently affect TM helical self-association in a family of epidermal growth factor receptors (96). Given the diverse and complex nature of the lipid bilayer, elucidating relationships between lipid bilayer compositions with TM protein oligomerisation is difficult and far from complete.

1.3 Low complexity sequence scaffolds

A TM low complexity scaffold is composed from the bare essential amino acids required to assemble an α-helix and to satisfy the required hydrophobicity for spontaneous insertion. These two requirements are typically fulfilled by using either a polyleucine sequence, a heptad repeat sequence (composed from two different non-polar amino acids) or some other arrangement of two different non-polar amino acids. What follows is an account in the literature of how low complexity scaffolds have been used to investigate low complexity sequences containing a small-xxx-small motif, a polar residue, or a duet of polar residues and small-xxx-small motifs. Unlike 'top-down' investigations carried out on native sequences, the available studies are limited. A significant number of low complexity sequences were constructed for single-site mutagenesis studies derived from the results of combinatorial plasmid libraries, or iterative reconstruction of a native protein based on heuristics from structural analysis. The following low complexity sequence review is separated according to the adjustment made to the low complexity scaffold, notably, GG4 or a small-xxx-small motif, polar or ionic residues, aromatic residues, and all other studies. We are particularly interested in the arrangement of the low complexity sequence, the experimental procedure performed (in addition to the membrane-mimetic), and the conclusions drawn.

1.3.1 GxxxG motif low complexity sequences

A polyvaline and polymethionine low complexity scaffold was modified through a process of iterative site-directed mutagenesis to identify the amino acid motif mediating glycophorin A homodimerisation (24) (*from M13*). Neither polymethionine nor polyvaline basic sequences demonstrated significant dimerisation from ToxR analysis. A single substitution of methionine/valine for glycine had a negligible affect on TM domain association until a second glycine residue was positioned, three residues from the first resulting in a significant increase in homodimerisation. Isoleucine, leucine, and tyrosine residues were added at the respective position found at the GpA TM domain: ^{75}L^{76}Ixx^{79}G^{80}Vxx^{83}G^{84}Vxx^{87}T. These additional adjustments to the basic sequence failed to encourage further increase in self-association. These results suggest that in the context of a polymethionine and polyvaline low complexity scaffold, a GG4 motif stabilises TM domain oligomerisation. Further to this work, a minimised GpA dimerisation motif was systematically grafted onto a polyleucine low complexity scaffold (51) (*from GG*). Given the large increase in propensity for dimerisation after the GG4 motif was added and prior to the additional GpA residues, polyleucine proves to form a stable scaffold to form strong TM domain oligomers. In addition, polyleucine scaffolds were used to investigate the effects of helix-helix association whilst adjusting the length of the hydrophobic helix by adding or removing flanking leucine residues. The ideal TM domain length according to three different PAGE analysis techniques is 18 to 24 leucine residues. Both studies (24, 51) were performed *in vivo*. The propensity for specific oligomeric states (dimer and trimer) was not reported.

The low complexity scaffolds polyleucine and polyalanine (97), were randomised at *inter*helical packing sites using a library of plasmid constructs (*Randomised A16/A17 library*). Sequences with a high-affinity for dimerisation were isolated using the *in vivo* TOXCAT system (25). Over 80% of the identified systems contained a GG4 motif. As seen earlier by Herrmann *et al.,* (79), GG4 isolates were accompanied by flanking residues, notably valine, leucine, isoleucine and threonine on a polyleucine scaffold, and a strong propensity for an additional glycine along with either valine, leucine or isoleucine on a polyalanine scaffold. Interestingly, the emerging patterns in the high-affinity isolated polyleucine sequences are similar to the iconic GpA sequences stated earlier, conforming to either G[large]xxG[large]xx[small] or G[small]xxG[small]xxT. The frequently

occurring patterns would suggest that these positions and amino acid types are indicative of a favourable packing motif. Either [large] and [small] residues beside each glycine residue is not surprising assuming that in either case the side chain would not pack directly at the helix-helix interface, therefore, the size of the size chain would not affect association.

Using the *in vivo* POSSYCATT (27) assay to screen a plasmid library of TM protein sequences for high helix-helix affinity sequences (49), the over-represented FxxGxxxG sequence was identified and analysed further by grafting the sequence onto a polymethionine low complexity scaffold (*from L16*). The results from adjusting the position of the phenylalanine yielded a high propensity for oligomerisation providing the FxxGxxxG motif was preserved. The proximity to the helix-helix closest point of approach, in addition to the position of the phenylalanine on the helix may suggest that a GG4 mediated packing motif can be stabilised further by potential π aromatic interactions. The FxxGxxxG motif was grafted onto two further low complexity scaffolds, AZ2 and polyleucine. Although, both sequences demonstrate a poor propensity for oligomerisation, after systematically replacing three native residues with an FxxGxxxG motif the AZ2 construct demonstrated significantly higher levels of oligomerisation when grafted onto the polyleucine low complexity scaffold. This would suggest that the mode of oligomerisaiton of this motif was highly sequence dependent.

GxxxG and small-xxx-small motifs

Name*	**Sequence**	**Reference**
*Naming convention from literature		
M13	MMMMMMMMMMMM	Brosig and Langosch (24).
M12.1	MMMM**G**MMMMMMMM	
M12.2	MMMMMMMM**G**MMMM	
M11	MMMM**G**MMM**G**MMMM	
M9	M**I**MM**G**MMM**G**MMM**T**	
M8	**LI**MM**G**MMM**G**MMM**T**	
V13	VVVVVVVVVVVVV	
V11	VVVV**G**VVV**G**VVVV	
V9	V**I**VV**G**VVV**G**VVV**T**	
V8.1	**LI**VV**G**VVV**G**VVV**T**	
V6	**LI**VV**GA**VV**GA**VV**T**	

V7	**AA**VV**G**VVV**G**VVAA	
GG	LLLLLL**G**LLL**G**LLLLLL	Orzáez*et al* (51).
GVGV	LLLLLL**GV**LL**GV**LLLLLL	
IGVGV	LLL**I**LL**GV**LL**GV**LLLLLL	
GGT	LLLLLL**G**LLL**G**LLL**T**LLL	
GVGT	LLLLLL**GV**LL**G**LLL**T**LLL	
GGVT	LLLLLL**G**LLL**GV**LL**T**LLL	
LLGVT	LLLLLLLLLL**GV**LL**T**LLL	
GVGVT	LLLLLL**GV**LL**GV**LL**T**LLL	
IGVGVT	LLL**I**LL**GV**LL**GV**LL**T**LLL	
GGS	LLLLLL**G**LLL**G**LLL**S**LLL	
GGN	LLLLLL**G**LLL**G**LLL**N**LLL	
GVGVS	LLLLLL**GV**LL**GV**LL**S**LLL	
GVGVN	LLLLLL**GV**LL**GV**LL**N**LLL	
Randomised A16 library	AS**xx**AA**xx**AA**xx**AA**x**AILI	Russ and Engelman (97).
Randomised L16 library	LS**xx**LL**xx**LL**xx**LL**x**LILI	
L16	LLLLLLLLLLLLLLL	Unterreitmeier *et al* (49).
L16-L5F/L8G/L12G	LLLL**F**LL**G**LLL**G**LLLL	
AZ2	LLAALLALLAALLALL	
AZ2-L8G/L12G	LLAALLA**G**LAA**G**LALL	
AZ2-L5F/L8G	LLAA**F**LA**G**LAALLALL	
AZ2-L5F/L8G/L12A	LLAA**F**LA**G**LAAALALL	
AZ2-L5F/L8G/L12G	LLAA**F**LA**G**LAA**G**LALL	
M11	MMMM**G**MMM**G**MMMM	
MM-M1F	**F**MMM**G**MMM**G**MMMM	
MM-M2F	M**F**MM**G**MMM**G**MMMM	
MM-M3F	MM**F**M**G**MMM**G**MMMM	
MM-M4F	MMM**FG**MMM**G**MMMM	
MM-M5F	MMMM**F**MMM**G**MMMM	
MM-M6F	MMMM**GF**MM**G**MMMM	
MM-M7F	MMMM**G**M**F**M**G**MMMM	
MM-M8F	MMMM**G**MM**FG**MMMM	
MM-M9F	MMMM**G**MMM**F**MMMM	
MM-M10F	MMMM**G**MMM**GF**MMM	

MM-M11F	MMMM**G**MMM**G**M**F**MM
MM-M12F	MMMM**G**MMM**G**MM**F**M
MM-M13F	MMMM**G**MMM**G**MMM**F**
MM-M2W	M**W**MM**G**MMM**G**MMMM
MM-M2Y	M**Y**MM**G**MMM**G**MMMM

1.3.2 Polar and/or ionic side chain motif low complexity sequences

A systematic asparagine-scanning mutagenesis of a polyleucine low complexity scaffold was performed and self-association was measured using the ToxR chimeric protein system (98) (*from Leu20*). The strongest impact on TM domain self-association was seen when asparagine was located around the centre of the TM domain sequence, at ^{7}L, ^{11}L, ^{14}L, and ^{18}L. These residues map to sites *d* and a on a heptad repeat sequence, placing them at the helix-helix interface of a leucine zipper. In addition, site-mutagenesis on heptad positions *g* and *e* at the centre of the TM domain, were seen to also stabilise oligomerisation suggesting that flanking asparagine side chains can stabilise a polyleucine low complexity TM domain. Oligomerisation was confirmed using SDS-PAGE analysis. The polyleucine control was shown to be monomeric, along with N-terminal asparagine site sequences L1N to L6N, and C-terminal asparagine site sequences L16 to L20. All other sequences were dimeric.

Position-specific histidine residues, flanked with glycine, serine and/or threonine, in addition to a C-terminal GG4 motif were selected using the TOXCAT assay from a combinatorial library of plasmid sequences on a polyleucine scaffold (50) (*randomised L16 library*). Mutational analysis on selected sequences confirmed that histidine side chains would form a network of *inter*helical hydrogen bonds, mediated via the tight packing of the GG4 motif.

*Inter*helical ionic interactions between oppositely charged side chains were shown to stabilise TM domain association depending on sequence context (79) from ToxR and site-directed mutagenesis analysis (*from L16*). In the presence of a C-terminal GG4 motif, a single charged aspartic acid moderately encouraged self-association with respect to a polyleucine sequence and a C-terminal GG4 motif. With the addition of a second oppositely charged ionic residue, arginine, TM domain

association is substantially increased. Finally, adding two additional residues, cysteine and tyrosine, or serine and histidine, increases association further from hydrogen bonding or an ionic-π interaction in the case of tyrosine interacting with arganine.

A low complexity scaffold of mixed leucine and alanine (*from 9TM+Leu*) was used to elicit the minimal dimerisation motif of the bacterial Tar-1 homodimer TM domain (99) (the arrangement of polar residues in Tar-1 can be seen in Figure 1.3 (B)). The polar residues, glutamine and serine when placed two residues apart, stabilised self-association. Interestingly, the QxxS motif grafted onto a polyalanine low complexity scaffold gave no indication of oligomerisation. As it would happen, although SDS-PAGE suggested that the polyalanine chimera proteins were being expressed, the absence of growth on a maltose minimal media plate suggested that polyalanine would not spontaneously insert into the bilayer. This may be due to the borderline hydrophilic nature of alanine in addition to the energetic penalty attributed to the cost of inserting serine and glutamine into a hydrophobic lipid environment.

A low complexity sequence of polyleucine residues (pVNVV and pVVVV) with an interfacial sequence of valine-asparagine-valine-valine was used to record free energy values before and after an asparagine to valine substitution (100). Computational potential of means force (PMF) calculations enables the decomposition of the free energy from helix-helix and helix-lipid interactions. The low complexity sequence provided an efficient platform to report significant per-reside energetic contributions, as is the case of a centrally located asparagine residue which suggests favourable patterns of *inter*helical hydrogen bonding between asparagine on a polyleucine scaffold.

Finally, inspired by a polyleucine heptad repeat with interfacial residues asparagine and valine, Zhou *et al.,* (70) positioned a single polar residue in interfacial sites (*a* and *d*) on a polyleucine low complexity sequence (*from L23*). Polar residues serine, threonine and tyrosine failed to encourage self association compared to the low complexity polyleucine control, possibly due to the propensity for these residues to form hydrogen bonds with the backbone oxygen. Whereas, asparagine, aspartic acid, glutamine, glutamic acid and histidine, residues were capable of being simultaneously hydrogen bond donors and acceptors.

Although, far from being a combinatorial analysis over all naturally occurring amino acids on a low complexity scaffold, Shi *et al.,* (101) used two polyalanine low complexity scaffolds to study

the potential interactions between a phenylalanine and glutamic acid (*from EFE4-4 and EFE505*). Their study is an indispensable example of low complexity sequences analysed using ^{1}H NMR spectroscopy, in sight of many *in vivo* helix-helix interactions assays reviewed thus far. They concluded that the hydrogen bond between the glutamic acid hydrogen and the phenylalanine oxygen, C-H$\cdots$O respectively, contributed approximately -0.5 kcal mol^{-1} per pair.

Low complexity scaffolds can be used to study other TM domain associated behaviour besides helix-helix packing. For example, out current understanding of how a sequence containing polar amino acids are translocated from the Sec61/SecY translocon into membrane bilayer without the assistance of existing helical TM proteins is limited. Using a TM domain sequence of repeating alanine-leucine pairs, a central residue was substituted with an arginine, a lysine, a glutamic acid, or an aspartic acid. MD simulations of low complexity TM helices were used as a basis for understanding how single charged amino acids behaved within the lipid core (102) (*from K14-D14*). Studies were performed by inserting pairs of low complexity scaffolds containing an oppositely charged central amino acid substitution. Several water molecules were shown to penetrate the bilayer and stabilise the charged side chains in addition to *inter*helical salt-bridge interactions between the charged side chains. A hydrogen bond network between permeated water and polar side chains could negate the energy penalty for inserting polar side chains into the non-polar lipid environment.

Polar and/or ionic motifs		
Name*	**Sequence**	**Reference**
*Naming convention from literature		
Leu20	LLLLLLLLLLLLLLLLLLL	Ruan *et al* (98).
L1N	**N**LLLLLLLLLLLLLLLLLL	
L2N	L**N**LLLLLLLLLLLLLLLLL	
L$\cdots$N	$\cdots$	
L19N	LLLLLLLLLLLLLLLLLL**N**L	
L20N	LLLLLLLLLLLLLLLLLLL**N**	
Randomised L16 library	L**x**LL**xx**L**xx**LL**xx**LL**x**	Herrmann *et al* (50).
L16	LLLLLLLLLLLLLLLL**G**	Herrmann *et al* (79).
L16D5	LLLL**D**LLLLLLLLLL**G**	
L16R6	LLLLL**R**LLLLLLLLL**G**	
L16D5R6	LLLL**DR**LLLLLLLLL**G**	

L16G13	LLLLLLLLLLLL**G**LLL**G**	
L16D5G13	LLLL**D**LLLLLLL**G**LLL**G**	
L16R6G13	LLLLL**R**LLLLLL**G**LLL**G**	
L16D5R6G13	LLLL**DR**LLLLLL**G**LLL**G**	
L16D5C8G13	LLLL**D**LL**C**LLLL**G**LLL**G**	
L16D5Y9G13	LLLL**D**LLL**Y**LLL**G**LLL**G**	
L16R6C8G13	LLLLL**R**L**C**LLLL**G**LLL**G**	
L16R6Y9G13	LLLLL**R**LL**Y**LLL**G**LLL**G**	
L16D5R6C8G13	LLLL**DR**L**C**LLLL**G**LLL**G**	
L16D5R6Y9G13	LLLL**DR**LL**Y**LLL**G**LLL**G**	
L16D5R6C8Y9G13	LLLL**DR**L**CY**LLL**G**LLL**G**	
L16R6	LLLLL**R**LLLLLLLLL**G**	
L16E8	LLLLLLL**E**LLLLLLLL**G**	
L16R6E8	LLLLL**R**L**E**LLLLLLLL**G**	
L16G13	LLLLLLLLLLLL**G**LLL**G**	
L16R6G13	LLLLL**R**LLLLLL**G**LLL**G**	
L16E8G13	LLLLLLL**E**LLLL**G**LLL**G**	
L16R6E8G13	LLLLL**R**L**E**LLLL**G**LLL**G**	
L16R6E8G9G13	LLLLL**R**L**E**G**L**LL**G**LLL**G**	
L16S5R6E8G9G13	LLLL**SR**L**E**G**L**LL**G**LLL**G**	
L16R6E8G9H12G13	LLLLL**R**L**E**G**L**L**H**G**L**LL**G**	
L16S5R6E8G9H12G13	LLLL**SR**L**E**G**L**L**H**G**L**LL**G**	
9TM+Leu	LLLLLLLLL**Q**LI**S**GSL	Sal-man *et al* (99).
QXXS+backbone	LLLLLLLAA**Q**AA**S**AAA	
Leu+Ala	LLLLLLLAAAAAAAAAA	
QXXS+Ala	AAAAAAAAA**Q**AA**S**AAA	
pVNVV	LLLLVLLLLLL**N**LLLLLLVLLLLLLVL	Lee and Im (100).
pVVVV	LLLLVLLLLLLLLLLLL**V**LLLLLLVL	
L23	RLLLLLLLLLLLLLLLLLLLLLLRRLI	Zhou *et al* (70).
VNVV	RVLLLLLL**N**LLLLLLVLLLLLLVLRRLI	
L7NL15	RLLLLLLL**N**LLLLLLLLLLLLLLLRRLI	
L7DL15	RLLLLLLL**D**LLLLLLLLLLLLLLLRRLI	
L7QL15	RLLLLLLL**Q**LLLLLLLLLLLLLLLRRLI	
L7EL15	RLLLLLLL**E**LLLLLLLLLLLLLLLRRLI	

L7HL15	RLLLLLLL**H**LLLLLLLLLLLLLLRRLI	
L7SL15	RLLLLLLL**S**LLLLLLLLLLLLLLRRLI	
L7YL15	RLLLLLLL**Y**LLLLLLLLLLLLLLRRLI	
L7TL15	RLLLLLLL**T**LLLLLLLLLLLLLLRRLI	
L19	RARLLLLLLLLLLLLLLLLLGILIN	
L7NL11	RARLLLLLLL**N**LLLLLLLLLLGILIN	
L7DL11	RARLLLLLLL**D**LLLLLLLLLLGILIN	
L7EL11	RARLLLLLLL**E**LLLLLLLLLLGILIN	
L7HL11	RARLLLLLLL**H**LLLLLLLLLLGILIN	
L7SL11	RARLLLLLLL**S**LLLLLLLLLLGILIN	
L7YL11	RARLLLLLLL**Y**LLLLLLLLLLGILIN	
L7TL11	RARLLLLLLL**T**LLLLLLLLLLGILIN	
EFE4-4	AAAAA**E**AAA**F**AAA**E**AAAAA	Shi *et al* (101).
EFE5-5	AAAA**E**AAAA**F**AAAA**E**AAAA	
K14-D14	GGPGAAAALALAL**K**LALALAAAAGPGG	Jusoh and Helms (102).
	GGPGAAAALALAL**D**LALALAAAAGPGG	
R14-D14	GGPGAAAALALAL**R**LALALAAAAGPGG	
	GGPGAAAALALAL**D**LALALAAAAGPGG	
K14-E14	GGPGAAAALALAL**K**LALALAAAAGPGG	
	GGPGAAAALALAL**E**LALALAAAAGPGG	
R14-E14	GGPGAAAALALAL**R**LALALAAAAGPGG	
	GGPGAAAALALAL**E**LALALAAAAGPGG	
K12-E16	GGPGAAAALAL**K**LALALALAAAAGPGG	
	GGPGAAAALALALAL**E**LALAAAAGPGG	
R12-E16	GGPGAAAALAL**R**LALALALAAAAGPGG	
	GGPGAAAALALALAL**E**LALAAAAGPGG	
R12-D16	GGPGAAAALAL**R**LALALALAAAAGPGG	
	GGPGAAAALALALAL**D**LALAAAAGPGG	
K12-D16	GGPGAAAALAL**K**LALALALAAAAGPGG	
	GGPGAAAALALALAL**D**LALAAAAGPGG	

1.3.3 Aromatic motif low complexity sequences

The electrostatic attraction between a positive charge and the conjugated π electrons of an aromatic ring was shown to mediate the self-association of a low complexity TM domain sequence of repeating alanine and isoleucine residues (103) (*From AAA*). Combinations of cation and/or aromatics pairings were positioned one residue apart. According to energy minimisation of a molecular model of the WAK TM protein sequence (AIAIAIIA**W**A**K**AIIAIAIAI), the arrangement of flanking residues resulted in the aromatic ring of tryptophane coming within interaction distance of the charged side chain of lysine on the neighbouring helix. Their propensity for self-association *in vivo* was assessed using a chloramphenicol acetyltransferase enzyme-linked immunosorbent assay. Results confirmed that although the principal van der Waals packing mediated self-association, the addition of electrostatic interactions between conjugated π and cation interactions were prominent. Interestingly, self-association of low complexity sequences with a single cation residue was shown to increase. Given that one would expect repulsion between charged side chains, if the pK_a was shifted, these residues would become deprotonated, resulting in *inter*helical hydrogen bond formation between side chains.

A randomised library of AZ2 heptad repeat TM domain sequences was screened using the POSSYCCAT system (104). A comparison between high dimer affinity sequences revealed an abundance of tryptophan. Against an AZ2 low complexity background, tryptophan placed at heptad sites *g* demonstrated a significant increase in self-association. Identically placed tryptophan on a polyleucine low complexity scaffold performed only half as well. This suggests that although tryptophan can encourage self-association in the context of both low complexity scaffolds, the difference between the scaffolds affects oligomerisation. It would have to be assumed that the same substitution on two low complexity scaffolds, which by themselves are equally poor dimers, suggests a sequence context driven by tryptophan association.

Low complexity scaffolds of mixed alanine-leucine and a polyleucine sequence had previously been used to investigate the interaction motif QxxS (99). This study was extended further by replacing the glutamine and serine residue with tryptophan (105) (*From Leu+Ala*). An increased level of oligomerisation was detected using the TOXCAT reporter assay. A planar arrangement of two aromatic rings would result in favourable π-π interactions, stabilising dimerisation.

Aromatic side chain motifs		
Name*	**Sequence**	**Reference**
*Naming convention from literature		
AAA	AIAIAIIAAAAAIIAIAIAI	Johnson *et al* (103).
WAA	AIAIAIIA**W**AAAIIAIAIAI	
WAK	AIAIAIIA**W**A**K**AIIAIAIAI	
WAH	AIAIAIIA**W**A**H**AIIAIAIAI	
WAR	AIAIAIIA**W**A**R**AIIAIAIAI	
YAA	AIAIAIIA**Y**AAAIIAIAIAI	
YAK	AIAIAIIA**Y**A**K**AIIAIAIAI	
FAA	AIAIAIIA**F**AAAIIAIAIAI	
FAK	AIAIAIIA**F**A**K**AIIAIAIAI	
AAH	AIAIAIIAAA**H**AIIAIAIAI	
AAR	AIAIAIIAAA**R**AIIAIAIAI	
AAK	AIAIAIIAAA**K**AIIAIAIAI	
AZ2	LLAALLALLAALLALL	Ridder *et al* (104).
AZ2-Wa	L**W**AALLAL**W**AALLAL**W**	
AZ2-Wd	LLAA**W**LALLAA**W**LALL	
AZ2-We	LLAAL**W**ALLAAL**W**ALL	
AZ2-Wg	**W**LAALLA**W**LAALLA**W**L	
L16	LLLLLLLLLLLLLLL	
L16-Wg	**W**LLLLLL**W**LLLLLL**W**L	
Leu+Ala	LLLLLLLAAAAAAAAA	Sal-man *et al* (105).
QXXS+backbone	LLLLLLLAA**Q**AA**S**AAA	
WXXW+backbone	LLLLLLLAA**W**AA**W**AAA	
Ala16	AAAAAAAAAAAAAAAA	
WXXW+Ala16	AAAAAAAAA**W**AA**W**AAA	

1.3.4 Miscellaneous low complexity sequences

Besides their use as a scaffold for site-mutagenesis, low complexity sequences have been used to understand the effects detergent, used to solubilise TM proteins, have on α-helical TM peptide

stability. Tulumello and Deber (80) sought to report the effects water, anionic, zwitterionic, neutral, and 'natural' (varying acyl tail length) detergents, had on the structure of TM domains of lactose permease, sodium/galactose symporter, vitamin B_{12} transporter and the arginine/agmatine antiporter. Two low complexity sequences were used to augment the studies as controls (*AI10 and AI5*). By adjusting the ratio of isoleucine to alanine residues, the authors were able to control the propensity for spontaneous insertion into the detergent. Both artificial peptides had a statistically equivalent level of helicity in all detergents according to Circular dichroism spectra. The monitored blue shift in tryptophan fluorescence spectroscopy was consistent with both peptides interacting with detergent micelles. In particular, AI10 demonstrated a much greater blue shift, indicative of the increase in hydrophobicity due to the number of isoleucine residues. The authors concluded that the denaturing properties of 'harsh' and 'mild' detergents is largely due to the interaction between detergent and non-membranous regions of proteins.

The propensity for self-association of the heptad repeat low complexity sequences polyleucine and AZ2 yielded statistically similar propensities for self-association within biological membranes (106) (*from L16*). Adjustments to the underlying AZ2 heptad sequence were made by firstly replacing the interfacial residues (heptad sites *a* and *d*) and then in a separate sequence replacing the side-interfacial residues (heptad sites *e* and *g*). Replacing either pair of leucine residues with alanine resulted in a significant drop in self-association. They assumed that the polyleucine and AZ2 sequences formed α-helical bundles upon self-association, stabilised via a 'knobs-into-holes' type of side chain packing due to the flexibility of the leucine side chain. The back facing heptad sites of AZ2 were replaced with alanine, and yet the propensity for self-association did not change. It is only when leucine residues at interfacial or side-interfacial sites are replaced by the less flexible alanine is there a reported drop in self-association.

Interestingly, one of the few *de novo* driven designs based on a low complexity scaffold was an investigation into TM domain flexibility for fusion protein complexes (107). Valine residues were systematically added to a polyleucine low complexity scaffold. The ratio of leucine to valine was crucial to ensure the ability to continually refold between an α-helical state to a β-sheet conformation. Valines at sites i and $i + 4$ may result in steric repulsion due its bulky side chain, which would decrease the number of residues per turn in a helix. Further still, a leucine residue was replaced with a helix-breaking proline and a helix-kinking glycine with the aim of increasing

the flexibility of the structure. The final construct demonstrated the required flexibility between conformations compared with a polyleucine TM domain.

Name*	**Sequence**	**Reference**
*Naming convention from literature		
AI10	AIAIAIIAWAIAIIAIAIAI	Tulumello and Deber (80).
AI5	AAAIAAIAWAIAAIAAAIAA	
L16	LLLLLLLLLLLLLLL	Gurezka *et al* (106).
A16	AAAAAAAAAAAAAAAA	
AZ2	LLAALLALLAALLALL	
L2/5/9/12A	LAAAALALAAAALALL	
L6/8/13/15A	LLAALAAALAALAAAL	
L16	KKKWLLLLLLLLLLLLLLLKKK	Hofmann *et al* (107).
LLV16	KKKWLLVLLVLLVLLVLLVLKKK	
LV16	KKKWLVLVLVLVLVLVLVLVKKK	
VVL16	KKKWVVLVVLVVLVVLVVLVKKK	
V16	KKKWVVVVVVVVVVVVVVVVKKK	
LV16–G8	KKKWLVLVLVLGLVLVLVLVKKK	
LV16–P8	KKKWLVLVLVLPLVLVLVLVKKK	
LV16–G8P9	KKKWLVLVLVLGPVLVLVLVKKK	
LV16–P8G9	KKKWLVLVLVLPGVLVLVLVKKK	

1.4 *De novo* TM peptide applications

De novo protein design aims to facilitate through rational protein engineering and synthetic biology ways of improving our understanding, knowledge and practical application of bionanotechnology and the manufacturing of biomaterials. Unfortunately, the TM protein-folding problem continues to evade all attempts at mastering the relationship between the peptide sequence and structure. Numerous efforts in solving the protein folding problem to improve our knowledge in both predictive and synthetic biology (108, 109) have brought about the idea of a rational protein design starting from seemingly low complexity peptide scaffolds. It is of no surprise that novel applications of *de*

novo engineered TM proteins are scarce at best.

Unlike the successful efforts made at functionalising *de novo* engineered soluble proteins, designing a TM protein with a specific application in mind remains elusive. It is clear from the limited literature available, that the number of successful *de novo* soluble protein designs, eclipses the successful efforts at developing *de novo* TM proteins. However, there have been some breakthroughs in early conceptual designs of electron TM electron transfer, essential for photosynthetic devices, and mimetic antimicrobial peptides. DeGrado and Lear (5) developed one of the first synthetic TM ion channels from a bundle of four α-helical peptides. Each helix was made from a repeat of leucine and serine to ensure one face was polar, and the other non-polar. This was later followed by advances in heme-bound *de novo* TM peptides. Four TM α-helices of an alanine-coiled motif were arranged into a bundle with two non-natural iron dipherylporphorins within the protein complex (110). The *de novo* design provides a multicentered pathway for transmembrane electron transfer. This was similar to an earlier study by Cordova *et al.,* (111) who synthesised the GxxxG containing GpA TM protein domain with a bis-His heme-binding site. The *de novo* TM dimer could catalyse the hydrogen peroxide dependent coordination of organic substrates such as 2,2',5,5'-tetramethyl-benzidine. Further work, investigating the same minimalist heme-binding membrane protein (112) as Cordova *et al.,* discovered that phenylalanine increased the binding affinity of iron by ten fold.

De novo design of membrane associating proteins has not just been limited to electron transfer complexes, in fact, the *de novo* design approach has been used to realize mimetic antimicrobial peptides. In particular, Signarvic and DeGrado designed switch-activated α-helical amphiphilic cell-lytic peptides based on the mastoparan lytic peptide family found in wasp venom (113). The synthesised peptides were engineered to bind to divalent cations zinc and nickel, resulting in a stable α-helical structure, which could then associate and disrupt the cell membrane. The apo-protein was unlikely to form a stable α-helical conformation. Therefore, the metal ion acts as a switch by interacting with two histidine residue or a histidine and glutamic acid residue.

The synthesis and purification of antimicrobial polymers with the intention of alternative antibacterial therapeutics is difficult and expensive. Tew *et al.,* (114) designed a series of *de novo* α-helical facially amphiphilic polymers which are easy to prepare from inexpensive monomers. Further to these designs, Tew *et al.,* (115) sought to design small synthetic oligomers which would adopt amphiphilic conformations and exhibit potent antimicrobial activity whilst remaining non-toxic to

host cells.

Very few successful efforts at functionalising TM α-helices into logic-gated mechanisms have made their way into literature. Most successful implementations of synthetic biological logic gates have used DNA bases as building blocks (116), or photoinduced electron transfer molecular structures (117). The nearest peptide-based logic gate, an auto-inhibited TM coiled-coil dimer switch, becomes active upon proteolytic cleavage resulting in the complementation of split-protein reporters on the N-terminal domain (118). This can provide for a dual protease sensor geometry that allows for the use of 'turn-on' sensors and as potential as an AND logic gate. The recent work by de Ronde *et al.,* (119) systematically explored dimer-ligand equilibria using a statistical mechanics model to realize sixteen possible logic gates. With the appropriate designed α-helical primary sequences, such a device could form the basis of a very basic protein-derived biochemical computer.

1.5 Aims and objections

The aim of this study is two fold. First, to elicit differences in free energy as a function of TM helix self-association between the oncogenic Neu receptor tyrosine kinase and its proto-oncogenic native form. Observing the effects of single point mutations in the context of a native sequence is a 'top-down' rationale. The polar and small-xxx-small motif in addition to the hypothesised means of activation of the Neu native protein, provides the motivation for our low complexity designs.

Secondly, to elicit free energy contributions to TM helix self-association of low complexity scaffolds before and after adding interaction motifs. One can consider these interactions as a contribution towards the 'rules of engagement' behind the rationale of *de novo* TM protein design, with the wider goal of designing purpose built protein systems in addition to further understanding the structure and folding of native and disease association TM proteins alike. The energetic contributions made by *inter*helical interaction motifs on *in vivo* low complexity heptad repeat sequences have yet to be studied in a natural membrane. In addition, whether a particular predetermined face for self-association of a low complexity heptad repeat sequence can be driven to switch upon the insertion of an interaction motif has yet to be investigated. In order to achieve this aim, this study is divided into two objectives:

1. Computational simulations and free energy calculations are used to investigate the effects of long range, lipid-mediated, helix-helix interactions during self-association of the Neu receptor tyrosine kinase.

2. Free energy values of self-association of a series of low complexity TM protein scaffolds are determined using the *in vivo* GALLEX assay and PMF calculations from coarse grained MD. Helix-helix GG4 and polar interaction motifs are substituted into low complexity scaffolds to measure the change in free energy, finishing with two sequences where a GG4 motif competes as a primary binding site for TM domain self-association. The set of TM domain sequences studied are presented in Figure 1.7. A flow diagram of energetic contributions and structural properties after including interaction motifs are recorded in the discussion of chapters 5 and 7.

Name	**Low complexity TM domain sequence**																
L17	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
AZ2	A	L	L	A	A	L	L	A	L	L	A	A	L	L	A	L	L
L17	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
L17(GG4)	L	L	L	L	L	L	G	L	L	L	G	L	L	L	L	L	L
L17(GG4)G11I	L	L	L	L	L	L	G	L	L	L	I	L	L	L	L	L	L
L17L9Q	L	L	L	L	L	L	L	L	Q	L	L	L	L	L	L	L	L
AZ2	A	L	L	A	A	L	L	A	L	L	A	A	L	L	A	L	L
AZ2L6S	A	L	L	A	A	S	L	A	L	L	A	A	L	L	A	L	L
AZ2L10S	A	L	L	A	A	L	L	A	L	S	A	A	L	L	A	L	L
$AZ2(GG4)_L$	A	L	L	A	A	G	L	A	L	G	A	A	L	L	A	L	L
$AZ2(GG4)_LG6I$	A	L	L	A	A	I	L	A	L	G	A	A	L	L	A	L	L
$AZ2(GG4)_LG10I$	A	L	L	A	A	G	L	A	L	I	A	A	L	L	A	L	L
$AZ2(GG4)_A$	A	L	L	A	A	L	L	G	L	L	A	G	L	L	A	L	L
$AZ2(GG4)_AG8I$	A	L	L	A	A	L	L	I	L	L	A	G	L	L	A	L	L
$AZ2(GG4)_AG12I$	A	L	L	A	A	L	L	G	L	L	A	I	L	L	A	L	L
$AZ2(GG4)_AL6A$	A	L	L	A	A	A	L	G	L	L	A	G	L	L	A	L	L
$AZ2(GG4)_AL10A$	A	L	L	A	A	L	L	G	L	A	A	G	L	L	A	L	L

Table 1.7: TM domain sequences of low complexity scaffolds with additional motifs. Motifs on low complexity sequences are in red.

1.6 Chapter Organisation

In chapter 2, the simulation theory is outlined. The experimental methodologies are reported separately in chapter 3. Due to the multi disciplinary nature of this study it was logical to document the relevant MD simulation setups in a separate section at the start of each simulation results chapter. The results of this study are divided into four chapters. The Neu receptor tyrosine kinase TM dimer is presented in chapter 4 as part of a 'top-down' investigation of lipid mediated TM oligomerisation

using MD. The alternative 'bottom-up' design is presented in chapter 5 using *in vivo* experiments to quantify the free energy of association of *de novo* TM proteins. Before complementing the experimental results with MD simulations, in chapter 6, we take a short detour to investigate and explain differences in computational free energy calculation presented in this study. This chapter acts as a bridge between the variation in computational techniques between chapter 4 and chapter 7. Finally, the results in chapter 7 complements the experimental work with MD simulations. Each of the result chapters begin with a self-contained introduction, methodologies, results and finishes with a discussion and conclusion. Chapter 8 brings this study to a close with a discussion on the work presented here and of the future direction of this field.

Chapter 2

Simulation Theory

2.1 Statistical mechanics

Many experimental techniques aim to yield thermodynamic and structural macroscopic bulk properties of a system, time averaged over a large number of particles. For example, Isothermal Titration Calorimetry (ITC) can return a number of thermodynamic and stoichiometric quantitative properties from the binding of molecules (120). In addition, techniques such as Circular Dichroism (CD) are able to approximate the average secondary structure composition due to the chiral nature of a protein's backbone (121). Classical computational simulations on the other hand, calculate the instantaneous position and velocity of atoms. Unfortunately, this kind of information cannot be compared directly to the averaged bulk properties from experimental data. It is therefore necessary to turn to statistical mechanics, which provides a framework for relating microscopic properties of individual atoms to macroscopic bulk properties such as work, heat, free energy, and entropy.

Statistical mechanics provides a framework to study macroscopic properties of a system of particles of the order 10^{23} by applying probability theory. Although typically today's computational simulations are only of the order 10^6 particles. We begin with the idea of a collection of microstates (system configurations) called an ensemble in a $6N$ dimensional *phase space*, Ω, where N is the number of atoms in the system. Although there are a number of ensembles which define a phase space, we are only interested in the canonical ensemble (*N, V, T*): constant number of atoms, system volume and temperature, and the isothermal-isobaric ensemble (*N, P, T*): constant

number of atoms, pressure and temperature. Each microstate is uniquely identifiable by the set $6N = (r_{x,1}, r_{y,1}, r_{z,1}, ..., r_{x,N}, r_{y,N}, r_{z,N}, p_{x,1}, p_{y,1}, p_{z,1}..., p_{x,N}, p_{y,N}, p_{z,N})$, where r is position and p is momentum. The ensemble is distributed across phase space according to a probability density distribution $\rho_{\text{ensemble}}(\Omega)$. A microstate shares some identical macroscopic properties, depending on the thermodynamic ensemble, for example total energy. As an example, this could be thought of as an identical arrangement of gas particles on either side of a box. Their configuration is different yet their energy is the same. The probability of finding a system in a particular microstate is proportional to the Boltzmann factor:

$$e^{-H(\boldsymbol{r},\boldsymbol{p})\beta}, \tag{2.1}$$

where H is the Hamiltonian, which represents the total energy as a function of all the positions and momenta of the system. The Boltzmann factor, equation 2.1, is a distribution measure based on the total energy of a system. β is given by:

$$\beta \equiv \frac{1}{k_B T}, \tag{2.2}$$

where k_B is the Boltzmann constant and T is the temperature. The sum of all microstates (also known as the sum over states) is defined by the partition function Z:

$$Z = \sum_{i=1}^{n} e^{-\beta H_i(\boldsymbol{p},\boldsymbol{r})}. \tag{2.3}$$

where n is the total number of microstates. In real systems, the set of microstates becomes uncountable as the position and momentum variables of a particle vary continuously. Therefore, it becomes inaccurate to represent the partition function as a set of discrete terms. The sum is replaced by integrating over position and momentum:

$$Z = (h^{3N} N!)^{-1} \iint d\boldsymbol{p} d\boldsymbol{r} e^{-H(\boldsymbol{p},\boldsymbol{r})\beta}, \tag{2.4}$$

where N is the number of particles, h is Plank's constant, and the term $(h^{3N}N!)^{-1}$ is used to prevent

over-counting the number of microstates in a system of identical particles. The partition function is the connection between macroscopic thermodynamic properties and microstates. Normalising the microstates of a given Hamiltonian $H(\boldsymbol{p}, \boldsymbol{r})$ by the partition functions yields the Boltzmann probability distribution:

$$\rho(\boldsymbol{p}, \boldsymbol{r}) = \frac{e^{-H(\boldsymbol{p},\boldsymbol{r})\beta}}{Z}. \tag{2.5}$$

The Boltzmann distribution is a probability measure of finding the system in one of any number of microstates. How does a system whose phase space spans the momentum and position of a vast number of particles yield information on the average of some property, A? Unfortunately, even for small systems, computing the partition function Z is impractical. It is understood that the average of A across the ensemble, known as the ensemble average, $\langle A \rangle_{ensemble}$, is proportional to the volume of phase space that a microstate of that particular system configuration will take up:

$$\langle A \rangle_{ensemble} = \iint d\boldsymbol{p}d\boldsymbol{r}A(\boldsymbol{r}, \boldsymbol{r})\frac{1}{Z}e^{-H(\boldsymbol{p},\boldsymbol{r})/k_BT}, \tag{2.6}$$

or using equation 2.5, it can be rewritten as:

$$\langle A \rangle_{ensemble} = \iint A(\boldsymbol{pq})\rho(\boldsymbol{p}, \boldsymbol{r})d\boldsymbol{p}d\boldsymbol{r}. \tag{2.7}$$

Assuming that the system is ergodic, where every microstate is accessible during the evolution of a system in time (see section 2.1.1), the ensemble average in equation 2.7 becomes a time average at the limit:

$$\langle A \rangle_{time} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\boldsymbol{p}, \boldsymbol{r})dt \approx \frac{1}{M} \sum_{t=1}^{M} A(\boldsymbol{p}, \boldsymbol{r}), \tag{2.8}$$

over M number of traversed points in phase space. The larger the value of M, the closer the time average quantity approximates the ensemble average property.

2.1.1 Ergodicity

As seen in equation 2.8, a system is said to be ergodic if the averaged behaviour over all microstates can be approximated by the averaged behaviour over time:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}. \tag{2.9}$$

The time spent by a system in some region of phase space is proportional to the volume of this region. Given current computational capabilities, the equality of time- and ensemble-averaged properties (equation 2.9) can only be validated for the simplest of systems. Rare events, such as the transition states between a reactant and a product occupy a very small portion of phase space.

Phase space can be seen as a carved up $6N$ dimensional landscape of energy barriers, energy minima, attractors, and valleys. The rate of crossing an energy barrier is proportional to the probability of finding the system at the top of the barrier. Fluctuations leading to exploration of an energy barrier are rare, yet the time required crossing the energy barrier is very short. Figure 2.1 is an illustration of an energy minima as a function of the distance between two TM α-helices. The area beneath the energy minimum (A) is significantly smaller than the area occupied by the helices acting as monomers (C), which would suggest that until the system experiences a descent (B), the configuration has a high likelihood of occupying anyone of the many microstates found in (C).

2.2 Interaction potentials

The time-evolution of a system of N particles is directed according to the forces experienced between particles. The forces between particles are calculated from a parameter set of non-bonded and bonded interaction potentials. Parameters will have been determined according to the philosophy of the force field, for example, based on *ab initio* calculations or experimental observations. The bonded and non-bonded interaction potentials plus the associated parameters, constitute what is known as a *force field*. The force fields used throughout this study are the Gromos53a6 (122) united atom and the MARTINI coarse grained (123). Coarse graining is summarised in section 2.3. The summation of bonded and non-bonded interactions for a single particle yields the total energy, of which the negative differential with respect to particle position returns the force on that particle. The

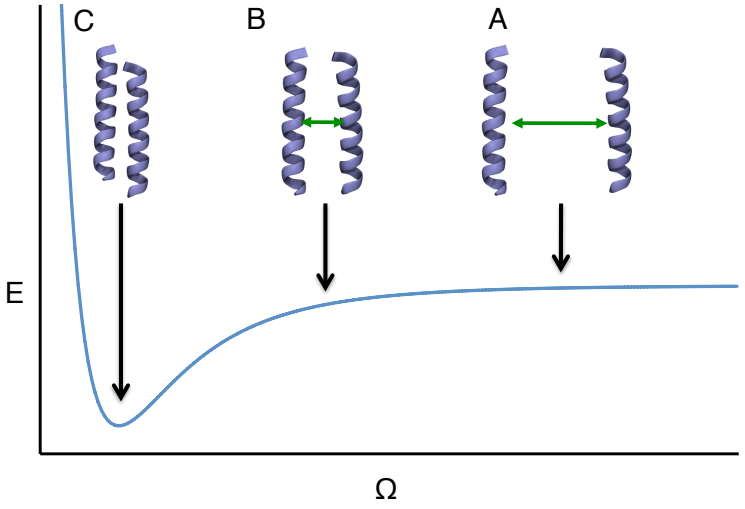

Figure 2.1: Rare events such as the oligomerisation of proteins (A through to C), traverse a potential energy pathway through the potential energy phase space (C and B). Under favourable conditions, the system descends into a free energy minimum upon TM protein dimerisation (A). The steep sides to the energy well prevent the system from escaping. The ergodic hypothesis states that the time spent in a region of phase space (Ω) is proportional to the volume occupied by microstates of the same energy

non-bonded interactions (see section 2.2.1 and 2.2.2) are a combination of three terms; repulsion, dispersion, and electrostatic or Coulomb. The repulsion and dispersion terms are combined to yield a Lennard-Jones (LJ) (124) or Buckingham (125) potential to name two. The electrostatic interactions between charged particles can be calculated using coulombs law; in practice this is achieved using an Ewald summation method (126) or a simple electrostatic interaction cut-off. Non-bonded interaction potentials are pair-additive between pairwise atoms i and j for a system of N atoms:

$$U(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \sum_{i<j} U_{ij}(\boldsymbol{r}_{ij}), \tag{2.10}$$

where U_{ij} is the non-bonded potential between two atoms as defined by their distance $\boldsymbol{r}_{ij}$, and according to Newton's third law, forces between pairwise atoms are symmetric:

$$\boldsymbol{F}_i = -\sum \frac{dU(\boldsymbol{r}_{ij})}{dr_{ij}} \frac{\boldsymbol{r}_{ij}}{r_{ij}} = -\boldsymbol{F}_j, \tag{2.11}$$

where $\boldsymbol{F}_i$ is the force on particle i, and $\boldsymbol{F}_j$ is the force on particle j.

2.2.1 Van der Waals interactions

Van der Waals interactions were modelled in this study using an LJ- potential. The general form of the LJ-potential is:

$$U_{LJ}(\boldsymbol{r}_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{\boldsymbol{r}_{ij}}\right)^{12} - \left(\frac{\sigma}{\boldsymbol{r}_{ij}}\right)^{6}\right], \tag{2.12}$$

where ϵ and σ describe the depth of the potential well, and the distance between pairwise particles ($\boldsymbol{r}_{ij}$) where the potential energy $U(r_{ij})$ is zero (see Figure 2.2). At the minimum of the energy well, denoted r_{m}, the potential energy equals $-\epsilon$. When derived, the LJ-potential term yields the contributing force with respect to the distance between particle i and j:

$$\frac{dU_{LJ}(\boldsymbol{r}_{ij})}{d\boldsymbol{r}_{ij}} = \boldsymbol{F}_i(\boldsymbol{r}_{ij}) = \left(12\frac{C_{ij}^{12}}{\boldsymbol{r}_{ij}^{13}} - 6\frac{C_{ij}^{6}}{\boldsymbol{r}_{ij}^{7}}\right). \tag{2.13}$$

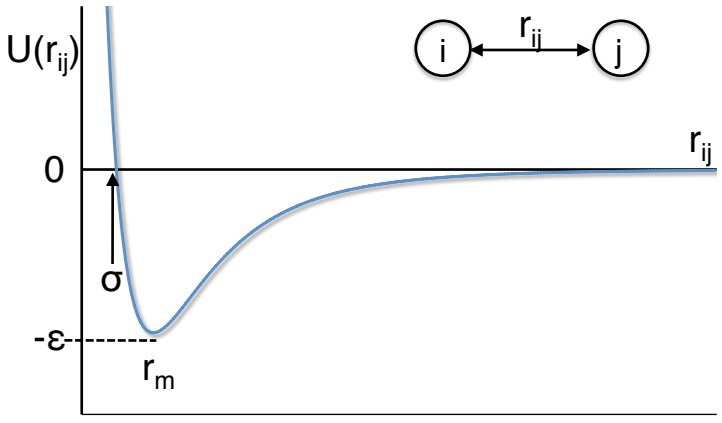

Figure 2.2: A representation of a LJ-potential between two atoms i and j, where ϵ and σ have been set to one.

The LJ-potential between particles runs asymptotically with interparticle distance. The LJ-potential can be modified using a truncated-shift cut-off, or a simple cut-off. The truncated-shift cut-off brings

the potential energy between particles to zero at a specified distance between atoms. Although, this technique helps conserve short-term total energy it can introduce artificially large forces. Alternatively, the LJ-potential can be truncated at a given distance without being shifted to zero. By doing so, the force calculations derived from the LJ-potential are improved at the expense of introducing a possible energy drift.

2.2.2 Electrostatic interactions

Long- and short-range electrostatic interactions can be modelled using the Particle Mesh Ewald (PME) summation (127, 128). Derived from the Ewald summation method (126), PME breaks the summed electrostatic interactions into a real space particle-particle short-range calculations and a Fourier space long-range calculation. The Fourier transformation of particle-particle interactions ensures efficient convergence of the algorithm across an infinite number of translated unit cells (see section 2.2.3.4 for unit cells and periodic boundary conditions). Unfortunately, the Ewald sum scales to $O(N^2)$. Yet, by distributing the charged points across a cardinal B-spline lattice, the PME algorithm is improved in efficiency by an order of $O(NlogN)$.

2.2.3 Covalent bonds

Covalent bonds contribute to the total energy of the system through the functional form:

$$U_{bonded,i} = U_b(r_{ij}) + U_a(\theta_{ijk}) + U_{id}(\xi_{ijkl}) + U_d(\theta_{ijkl}), \tag{2.14}$$

where $U_b(r_{ij})$ is the potential energy function for bond stretching between two atoms, $U_a(\theta_{ijk})$ is the potential energy function of the angle bending between three atoms, $U_{id}(\xi_{ijkl})$ is the potential energy function of keeping the correct chirality, and planar structure of carbon rings through the use of an improper dihedral, and $V_d(\theta_{ijkl})$ is the torsional energy associated with bond rotation. Each term is described in turn below.

2.2.3.1 Harmonic bond stretching

Bonds between two particles can be maintained using a harmonic bond potential of the form:

$$U_b(r_{ij}) = \frac{1}{2}k^b_{ij}(r_{ij} - b_{ij})^2, \tag{2.15}$$

where k_{ij} is a spring constant between particles i and j, and $(r_{ij} - b_{ij})^2$ is the squared distance from some ideal position b_{ij}. The force is given by the form:

$$F_i(r_{ij}) = k^b_{ij}(r_{ij} - b_{ij})r_{ij} \tag{2.16}$$

The harmonic bond stretching potential, although of interest to this study, is only one of many other bond stretching potentials. For example, the fourth power potential (122), which is typically computationally more efficient, is not of a harmonic form and thus the average energy of a single bond is not equal to $\frac{1}{2}kT$. Fourth power potentials are typically implemented for the GROMOS-87 forcefield (129). The Morse potential bond stretching (130) is typically used when an anharmonic bond stretching potential is necessary. Finally, a cubic bond stretching potential (131) is another anharmonic bond stretching potential, which is a simpler implementation than the Morse potential.

2.2.3.2 Harmonic angle potential

Much like the harmonic bond stretching potential, the harmonic angle potential between three bonded atoms, i, j, and k, uses a harmonic constant k^θ_{ijk} as a spring constant with an angular deviation θ_{ijk} about some idealised angle θ^0_{ijk}. The function takes the form:

$$U_a(\theta_{ijk}) = \frac{1}{2}k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2. \tag{2.17}$$

The force on each particle from the angle θ_{ijk} can be derived from the change in potential energy with respect to the particle's position:

$$\boldsymbol{F}_i = -\frac{dU_a(\theta_{ijk})}{d\boldsymbol{r}_i}, \tag{2.18}$$

$$\boldsymbol{F}_k = -\frac{dU_a(\theta_{ijk})}{d\boldsymbol{r}_k}, \tag{2.19}$$

$$\boldsymbol{F}_j = -\boldsymbol{F}_i - \boldsymbol{F}_k. \tag{2.20}$$

The cosine based angle potential is a simplified harmonic potential and is typically implemented for the GROMOS-96 force field. A harmonic correction term between the i and k particle as used for the CHARMM force field can be implemented using the Urey-Bradley potential (132). The bond-bond cross term and bond-angle cross term (133) are examples of non-harmonic angular potentials.

2.2.3.3 Harmonic improper dihedral potential

Improper dihedral angles are meant to ensure that planar groups, such as carbon rings on the side-chains of aromatic amino acids, remain planar and that ring structures do not flip. They are typically implemented using either a harmonic potential or a periodic potential. The harmonic potential takes the form:

$$U_{id}(\xi_{ijkl}) = \frac{1}{2}k_\xi(\xi_{ijkl} - \xi_0)^2. \tag{2.21}$$

where ξ_{ijkl} is the dihedral angle between atoms i, j, k and l, and ξ_0 is the ideal angle as defined in the force field.

2.2.3.4 Proper dihedral

A proper dihedral potential represents the torsion angle potential about four particles i, j, k, and l, resulting in the angle between the plane of particles i, j, and k with the plane of particles j, k, and l. An angle of zero corresponds to the *cis* configuration. The two most common implementations of a proper dihedral potential are either the periodic or Ryckaert-Bellemans (134) function. Each requires special consideration. The periodic angular potential requires a special 1-4 LJ-interaction between the first and fourth atom in a chain, whereas the Ryckaert-Bellemans function requires the exclusion of 1-4 non-bonded interactions for alkanes. In this study, a periodic function was implemented to model the dihedral angles of the lipid head group region, and Ryckaert-Bellemans function was implemented to model the dihedral angles of the lipid chains. The period function takes the form:

$$U_d(\theta_{ijkl}) = k_\theta(1 + \cos(n\theta - \theta_s)), \tag{2.22}$$

where k_θ is the force constant on the angle between the two planes, θ_s is the idealised dihedral angle defined in the force field, θ is the measured angle, and n is a multiplicity constant defined in the force field. The Ryckaert-Bellemans function takes the form:

$$U_{rb}(\theta_{ijkl}) = \sum_{n=0}^{5} C_n(cos(\psi))^n, \tag{2.23}$$

Where $\psi = \theta - 180°$. C is a constant, defined according to the philosophy of the force field. An alternative to a periodic or Ryckaert-Bellemans function, is a Fourier function (135).

2.3 Coarse grained force field

A coarse grained force field enables MD simulations to investigate system sizes inaccessible to all-atom or united-atom simulations. Essentially all forms of CG approaches group several atoms into new CG sites. This can be as few as several atoms, a whole or part of a side chain, or several amino acids. The CG quasi-particles interact through an effective potential (123), which will have been approximated using either an experimental/thermodynamic approach or by building on an existing atomic system. The reduction in the number of degrees of freedom and the combination of these effective potentials ensures a reduction in computational time by as much as three orders of magnitude compared with most atomistic simulations. As a result, CG computational approaches are seen as a bridge between atomistic and mesoscopic scales. For example, lipid flip-flopping (136), actin filament networks (137) and other large and complex biological phenomena which are typically outside of the scope of atomistic simulations.

There is a long history of CG molecular systems. As far back as 1975, the first CG protein simulation by Levitt and Warshel (138) sought to improve on computationally efficiency in biosimulations by only retaining the essential dynamics of protein folding. This was followed a few years later by the very influential CG Gō model (139), which yields a very smooth energy landscape, or rather a funnel pointing towards the native biomolecular structure. This is due to a Gō model only considering native contacts as being energetically favourable.

Typically, empirically based approaches may not be considered fully rigorous, however the approach can be improved by incorporating a statistical mechanics framework to add the benefits of a physics-based CG model. For example, CG approaches by Scheraga *et al.,* (140) and Klein *et al.,* (141) incorporate a blend of empirical CG modelling and physics-based modelling. The MARTINI force field, as used in the CG simulations of this study, is parameterised based on the reproduction of partitioning free energies, rather than reproducing structural details of particular systems. Although, this suggests that the model is empirical in nature, the parameterisation scheme of MARTINI has become valuable for membrane systems.

2.4 Potential energy minimisation

The previous section gave an account of the *inter*molecular functions, which are combined to form part of a force field. In this section we describe the concept of a potential energy minimisation algorithm. As a result of solvating a protein or bilayer, steric clashes between atoms or beads can be introduced during the initial creation of a system. The proximity between these problematic particles result in high intermolecular forces which can cause a system to 'explode'. To help prevent steric clashes, the potential energy between particles can be minimised through the use of one or more potential energy minimisation algorithms. The algorithms do not use the classical equations of motions to reach a minimum; rather, the algorithm gently perturbs the configuration for a predefined number of steps or until the maximum force in the system drops below a given threshold. Three of the most common potential energy minimisation algorithms are; steepest descent (142), conjugate gradients (143), and limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) quasi-Newtonian minimiser (144, 145). Only the steepest descent approach is reviewed here due to its use during this study.

2.4.1 Steepest descent

The steepest descent algorithm is a first order optimisation algorithm. Adjustments to a system of $3N$ coordinates follow steps proportional to the negative of the potential energy gradient. The algorithm begins with an initial maximum displacement h_0, for example, 0.01 nm. Forces and potential energy, U, are calculated at each step n.

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{max(|\mathbf{F}_n|)} h_n. \tag{2.24}$$

The notation $max(|\mathbf{F}_n|)$ refers to the largest of the absolute values of the force components. If $(U_{n+1} < U_n)$ the new positions are accepted and an adjustment is made to the displacement variable, for example, $h_{n+1} = 1.2h_n$. Otherwise, the new positions are rejected and a smaller adjustment is made to the displacement variable, for example, $h_n = 0.2h_n$.The steepest descent is limited by the asymptotic rate of convergence. Progression to the minima rapidly slows down, zig-zagging indefinitely.

2.5 Molecular Dynamics

Classical MD is viewed as the reduction from a fully quantum description to a classical potential. This is achieved through two approximations. The first, the Born-Oppenheimer approximation states that electrons react instantaneously to the motion of their nuclei; therefore they can be treated separately. The second approximation states that the nuclei are considered as point particles that follow classical dynamics. MD simulations allow the study of biological and chemical systems at the atomistic level on timescales from femto- to milli-second time scales. The simulation explores different structural conformations by making adjustments to the coordinates of each atom. These adjustments are the product of a time integration algorithm, which uses the summed derivation of the force field potential energy to compute the force on each particle. Unlike MC simulations, MD simulations are very similar to a real experiment. A system is prepared, left to equilibrate, and then allowed to propagate in time whilst taking a measurement. The basic MD algorithm is summerised in Table 2.1.

2.5.1 Equation of motion

In this section, we summarise the Newtonian equations of motion for a system of N particles interacting via a potential U. During a MD simulation new positions and velocities are calculating using Newton's second law:

$$\boldsymbol{F}_i = m_i \boldsymbol{a}_i = -\frac{dU_i}{d\boldsymbol{r}_i}. \tag{2.25}$$

1) Construct initial system coordinates. Assign random velocities from a Boltzmann distribution according to a given temperature. **2)** Calculate the force on all atom through the sum of non-bonded atom pairs: $\boldsymbol{F}_i = \sum_j \boldsymbol{F}_{ij}$, plus all forces due to bonded interactions, restraints and external forces. **3)** Compute the new position of all atoms after a small time step using Newton's equations of motion. **4)** Output positions, velocities, temperature, volume, pressure, and other macroscopic properties. Go to **2)**.

Table 2.1: The basic Molecular Dynamics algorithm. After the initial coordinates are set up, steps 2 to 4 loop until a given stop condition is satisfied.

Equation 2.25 states that the force **F** is exerted on atom i is equal to the product of the acceleration, **a**, and the mass, m, of that atom. Alternatively, the force can be expressed as the negative gradient (∇) of the potential energy:

$$\boldsymbol{F}_i = -\nabla U_i, \tag{2.26}$$

for $i = 1, 2, ...N$, and where the potential U_i is the summed contributions from non-bonded and bonded interactions, as seen in equations 2.10, 2.14. Combining equations 2.25, 2.26 yields:

$$-\frac{dU}{d\mathbf{r}_i} = m_i \frac{d^2\mathbf{r}_i}{dt^2}, \tag{2.27}$$

which relates the change in position of the atom with the derivative of the potential energy. The displacement of an atom position $\boldsymbol{r}(t)$ by a time step Δt requires knowledge of the velocity:

$$\boldsymbol{v} = \int_t^{t+\Delta t} \frac{d\boldsymbol{r}^2}{d^2t} = \int_t^{t+\Delta t} \frac{1}{m}\frac{d\boldsymbol{U}}{d\boldsymbol{r}} d\boldsymbol{r}. \tag{2.28}$$

As the force on an atom depends on the position of all other atoms in the system, it because clear that even for small systems there is no analytical solution to equation 2.28. Rather, a solution must be found numerically by utilising one of a number of time integration algorithms.

2.5.1.1 Time integration algorithms

Time integration algorithms are methods for numerically integrating Newton's second law of motion. It is important that algorithms preserve energy (symplectic), are time reversible, computationally efficient, and preserve the fastest vibrational frequency in the system by using an adequate time step. It is also important that integrating the equations of motion yields a true trajectory. Theoretically this is possible, however, due to the finite size of floating point arithmetic on classical computing devices there will always be some order of error. As such, the trajectory through phase space approximates the *true* trajectory (one which does not suffer from floating point arithmetic error) at best. It is not the aim of MD to predict system configurations, but rather to obtain the average behaviour. This is only possible providing sufficient effort has been invested in setting up the initial configuration. An additional feature important to time integration algorithms is that they preserve phase space volume. Typically, algorithms that do not preserve phase space end up increasing the phase space volume. As a consequence, a simulation could get trapped in a region of phase space as the neighbouring set of states continue to expand out as new states are introduced.

There are several algorithms equivalent to the Verlet approach. Of interest to this study due to the MD code implementation, is a second order method, both symplectic in nature and time-reversible, known as the leap-frog algorithm (146). To define the leap frog algorithm, velocities are calculated at half-integer time steps:

$$\boldsymbol{v}(t - \frac{\Delta t}{2}) = \frac{\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t)}{\Delta t}, \tag{2.29}$$

where $\boldsymbol{r}(t - \Delta t)$ is defined as

$$\boldsymbol{r}(t - \Delta t) = \boldsymbol{r}(t) - \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{F}(t)}{2m}\Delta t^2 - \dddot{\boldsymbol{r}} + O(\Delta t^4) \tag{2.30}$$

and

$$\boldsymbol{v}(t + \frac{\Delta t}{2}) = \frac{\boldsymbol{r}(t + \Delta t) - \boldsymbol{r}(t)}{\Delta t}, \tag{2.31}$$

where $\boldsymbol{r}(t + \Delta t)$ is defined as:

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{F}(t)}{2m}\Delta t^2 + \dddot{\boldsymbol{r}} + O(\Delta t^4). \tag{2.32}$$

From equation 2.31 we obtain an expression for the new position based on the old position and the velocities:

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \Delta t \boldsymbol{v}\left(\frac{t + \Delta t}{2}\right). \tag{2.33}$$

The update of the velocities occur at half-integer time steps:

$$\boldsymbol{v}(t + \frac{\Delta t}{2}) = \boldsymbol{v}(t - \frac{t\Delta}{2}) + \Delta t \frac{\boldsymbol{F}(t)}{m}. \tag{2.34}$$

Although the leap frog algorithm yields a trajectory identical to the Verlet scheme, the velocities are not defined at the same time as position, and therefore kinetic energy and potential energy are not calculated at the same time making it impossible to obtain the total energy of the system at the current time step.

2.5.2 Constant temperature - canonical ensemble

The microcanonical ensemble (N, V E); a fixed number of particles, constant system volume and constant volume, is far removed from simulating the behaviour of an experimental environment, where the temperature is typically controlled by the inclusion of a cellular environment, or the controlled temperature of a laboratory. Given that kinetic energy is the extensive counterpart to temperature, one could fix the total kinetic energy to approximate a specific temperature. However, this would not capture the true kinetic ensemble of the system. It would be experimentally more relevant to switch from the microcanonical ensemble to a canonical ensemble (N, V, T) by approximating an average temperature through the use of a thermostat. The temperature is allowed to fluctuate around an average by adding and removing energy to and from the system. A thermostat can also help avoid energy drifts caused by the accumulation of numerical errors during time integration.

There are several thermostats available including: velocity scaling (147); the Andersen thermostat (148); the Berendsen thermostat (149); the Nosé-Hoover thermostat (150); the Langevin (stochastic) thermostat (151); and the colored-noise Langevin thermostat (152). Below, we review the two thermostats used this study.

2.5.2.1 Berendsen thermostat

By coupling a thermal bath to the system, the velocities are gradually scaled proportional to the differences between the system temperature and the temperature of the thermal bath. Known as the Berendsen thermostat, it takes the form:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t)), \tag{2.35}$$

where τ is the strength of the coupling between the heat bath and the system, T_0 is the desired temperature and $T(t)$ is the temperature at time t. At the limit $\tau \to \infty$, the system approximates a microcanonical ensemble. This limit is never reached, and therefore the system never approximates a microcanonical ensemble. Yet, the Berendsen thermostat has the advantage of exponentially decaying towards an equilibrium without wildly oscillating, which makes it an ideal thermostat during early equilibration steps.

2.5.2.2 Nosé-Hoover thermostat

The Nosé-Hoover thermostat is implemented by extending the Langevin dynamics of the system by introducing an extra degree into the Hamiltonian of the system to take the form:

$$H(\boldsymbol{p}, \boldsymbol{r}, p_s, s) = \sum_i \frac{\boldsymbol{p}_i^2}{2ms^2} + \frac{1}{2}\sum_{i,j,i\neq j} U(\boldsymbol{r}_i - \boldsymbol{r}_j) + \frac{p_s^2}{2G} + gk_bTln(s), \tag{2.36}$$

where G is a constant which controls the coupling strength, p_s is the momentum of the extra degree of freedom, s, and g is the total number of degrees of freedom. The first two terms represent the kinetic energy and the potential energy. The third and forth terms are the kinetic and potential energy of the thermostat, respectively. Unlike the Berendsen thermostat, the Nosé-Hoover thermostat was shown to approximate a canonical ensemble, although the thermostat can fluctuate wildly for a system that is not in equilibrium (150). In addition, by replacing a coupled bath (that is to say,

the Berendsen thermostat) with an extra degree of freedom, the Nosé-Hoover thermostat has the advantage of being computationally inexpensive.

2.5.3 Constant pressure

The isothermal-isobaric (*N, P, T*) ensemble is a constant pressure extension to the canonical ensemble. Typically, this is achieved in one of two ways; a weak coupling to a pressure bath similar to the Berendsen thermostat, or an extension to the Hamiltonian by including an extra degree of freedom similar to the Nosè-Hoover thermostat.

2.5.3.1 Berendsen barostat

A Berendsen barostat (149) controls the pressure of a system by weakly coupling the system to an external pressure bath using the principal of least local perturbation. Coordinates and box vectors are scaled proportional to the compressibility of the system through exponential relaxation. The change in pressure over time is proportional to a diminishing approximation of the pressure P to a reference pressure P_0:

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \tag{2.37}$$

The Berendsen barostat yields a simulation with the correct average pressure, however as the limit $\tau_p \to \infty$ is never reached, the algorithm fails to ever yield an exact (*N, P, T*) ensemble (153) .

2.5.3.2 Parrinello-Rahman barostat

The Parrinello-Rahman barostat (153) uses an extended ensemble algorithm to allow the volume and shape of the cell to fluctuate. The hamiltonian is extended by including a thermal reservoir term s and a friction parameter ξ. The hamiltonian takes on the form:

$$H = K + V + K_s + V_s, \tag{2.38}$$

where K and V are the kinetic and potential energy terms, respectively, and the equation of motion becomes:

$$\frac{d^2\boldsymbol{r}_i}{dt^2} = \frac{m_i}{\boldsymbol{F}_i} - \xi\frac{d\boldsymbol{r}}{dt}. \tag{2.39}$$

The acceleration of an atom i is reduced by some factor $\xi\frac{d\boldsymbol{r}}{dt}$.

The Parinelo-Rahman barostat is similar in implementation by the addition of an extra degree of freedom to the Nosè-Hoover thermostat and as such if the system is far from equilibrium there is a tendency for the box the oscillate wildly. However, this algorithm does theoretically yield a simulation with an exact (*N, P, T*) ensemble.

2.5.4 Constraints

Ideally, the size of the time step used to evolve the Newtonian system should be small enough to capture all intramolecular atomic interactions, including the higher frequency intramolecular bond vibrations. Yet if this were the case, even the simplest of computer simulations would be excessively time-consuming. To resolve high-frequency vibrational motions, rather than reducing the time step bonded particles can be constrained, for example, by implementing the traditional SHAKE algorithm (154) or with the newer LInear Constraint Solver (LINCS) algorithm (155). Both algorithms make a correction to the set of new atomistic positions for all atoms connected by constrained bonds. During this study, all bonds were constrained using the Gromacs implementation of LINCS.

2.5.5 Periodic boundary conditions

Periodic boundary conditions yield a bulk phase by mimicking an infinite lattice of N identical particle unit cells. That is to say, a particle moving out of the unit cell is brought back through the opposite unit cell surface as if it were traveling through infinite space. The implementation of periodic boundaries results in what is known as a minimum image convention between pairwise particles. The nearest image of each particle is considered for short-range non-bonded interactions. This convention is not accurate enough for long range electrostatic interactions, and therefore the Ewald Sum lattice sum method is incorporated. A unit cell can take on the shape most suitable for the system under study. For example, the rhombic dodecahedron and the truncated octahedron (146) are closer to a sphere than a cube, and are therefore better suited for spherical macromolecules in

solution, since fewer solvent molecules are required to fill the primitive cell. A cubic unit cell suits the planar structure of a small region of cell membrane.

2.6 Free energy methods

The transition between two states governed by a difference in free energy provides insight into many molecular mechanisms, such as the folding of a protein, the diffusion of a molecule through a bilayer and the C-terminal activation of a cellular signalling cascade. Unfortunately, the calculation of thermodynamic properties, such as Gibbs free energy, are not directly accessible from MD simulations. However, there are at present a number of MD extensions that will yield thermodynamic properties. A themed issue of the Journal of Computational Chemistry in 2009 (156) provides an extensive review of some of the latest advancements in computational free energy techniques, theory and implementation, which are not surmised here.

Calculating the free energy and entropy between two states requires the complete consideration of all accessible microstates, their energy, and subsequently determining the partition function as defined in 2.4. For even some of the simplest of systems this is extremely impractical requiring hundreds if not thousands of computational hours. For full proteins and model membrane systems it is all but impossible with our current computational capabilities. Therefore, a number of free energy techniques have been developed and deployed with success.

Broadly speaking, free energy methods can be divided into two groups: equilibrium and non-equilibrium. Non-equilibrium methods rely on Jarzynski's equality (157) and have the potential to provide rapid free energy estimates but at the cost of introducing bias. This study only concerns itself with equilibrium techniques, where the approach relies on sufficient sampling to capture the system at a quasi-thermodynamic equilibrium.

To sample regions of high energy, the system can be restrained at a number of positions along a predefined path by applying a biasing potential. This is known as umbrella sampling (158). After sufficient sampling of overlapping regions of the predefined path, the biased distribution of samples can be unbiased using the Weighted Histogram Analysis Method (WHAM) (159). Differences in free energy can be calculated between states from the resultant potential of mean force profile.

There are equilibrium based free energy methods that give more freedom to explore the phase space

without predefining a set of configurations along a predefined pathway. These include adaptive biasing force (160) and metadynamics (161). Both techniques apply biasing potentials during the time evolution of the system along a number of predefined pathways e.g., angles, distances. This encourages the simulation to explore regions of phase space symptomatic of high energy barriers. Hénin *et al*., used adaptive biasing force calculations on the association of an all atom model of Glycophorin A in a palmitoyl-oleylphosphatidylcholine bilayer to yield the free energy contribution from a GxxxG motif (162). Their study was the first to report free energy calculations of the free energy profile associated with the reversible dissociation of a membrane protein. The free energy calculations presented in this thesis follow the same reversible dissociation reaction coordinate. Interestingly, they suggest the concept of retaining the main functionality of a dimer after a large fraction of the protein has been reduced to some low complexity scaffold. In essence, a large portion of the work presented in this thesis follows this concept.

2.6.1 Reaction coordinate

The free energy change between two meta-stable states, can be monitored as a function of a set of coordinates along a predefined path, better known as a reaction coordinate:

$$\xi(\boldsymbol{r}) \equiv \xi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{3N}). \tag{2.40}$$

This restricts the system to a region through phase space, making the free energy (Gibbs or Helmholtz) the partition function Z_ξ, and the reaction coordinate probability P_ξ, the relevant properties of the system to be determined. To begin with, the probability P of finding a system in a configuration defined by $(\mathbf{p}, \mathbf{r})$, at some point ξ^i along a reaction coordinate ξ is a simple extension to the Boltzmann probability distribution (equation 2.5):

$$P_\xi(\boldsymbol{p}, \boldsymbol{r}; \xi') = \rho(\boldsymbol{p}, \boldsymbol{r})\delta(\xi' - \xi(\boldsymbol{r})). \tag{2.41}$$

The partition function as defined in equation 2.4 is extended by included a Dirac delta function, δ, to ensure that only the distributions over the reaction coordinate are calculated:

$$Z_{\xi}(\xi_i) = (h^{3N}N!)^{-1} \iint d\boldsymbol{p}d\boldsymbol{r}\delta(\xi_i - \xi(\boldsymbol{r}))e^{-H(\boldsymbol{p},\boldsymbol{r})/\beta}. \tag{2.42}$$

By normalising the partition function (equation 2.42) by the unbiased partition function, we are able to calculate the probability of the system being in the state as defined by its position along the reaction coordinate:

$$\langle P_{\xi}(\xi_i)\rangle = \frac{Z_{\xi}(\xi_i)}{Z}. \tag{2.43}$$

2.6.2 Umbrella sampling

In chapters 4, 6 and 7, I present free energy profiles of helix-helix association. Interhelical distance is restrained at a specific position along the reaction coordinate using a biasing potential, enabling sufficient sampling of that region of phase space. These biased simulations are typically generated with an extension to the potential energy in the form $[U(\boldsymbol{r}) + w(\xi)]$. The simulations in this study have been restrained using the harmonic form $w_i(\xi) = \frac{1}{2}K(\xi_i - \xi)^2$, centred on successive values of ξ. After sufficient sampling, the sampling distribution approximates a gaussian distribution around some point ξ, hence the term 'Umbrella Sampling.' Biased distribution are obtained from the Boltzmann weighted average defined in equation 2.43:

$$\langle P_{\xi}(\xi_i)\rangle = \frac{\langle P(\xi)\rangle e^{-w_{\xi_i}/\beta}}{\langle e^{-w_{\xi_i}/\beta}\rangle}. \tag{2.44}$$

The results of each window are unbiased and recombined to yield the final estimate of $G(\xi)$. By reformulating equation 2.43 and using equation 2.44, the unbiased free energy at ξ_i becomes:

$$G_i(\xi) = G(\infty) - kTln\left[\frac{\langle P_i(\xi)\rangle}{Z}\right] - w_i(\xi) + F_i. \tag{2.45}$$

The unknown free energy constants F_i are obtained by adjusting each $G_i(\xi)$ until adjacent umbrella windows match. There are a number of algorithms readily available, one of which is covered in section 2.6.3.

2.6.3 Weighted Histogram Analysis Method

There are a handful of algorithms that aim to solve the PMF from a collection of umbrella distributions, for example, the least square estimation (163) approach and a novel combination of umbrella sampling and thermodynamic integration (164). Typically, older methods such as the least square estimation requires significant overlap between windows to overcome statistical error (163). The WHAM method (159), which is based on a generalised histogram method (165), uses the concept of maximum overlap of umbrella windows (166). It has the advantage of using all data points to reconstruct the unbiased distribution, however WHAM can be prone to give a poor estimate of the true PMF if the data points contributed to each histogram are not independent i.e., there should be no auto-correlation in time.

The WHAM works by constructing an estimate of the unbiased distribution function as a weighted sum over the number of data points, n, (equation 2.46) and constructs a functional of the weighted sum which minimises the statistical error. This begins with an initial estimate of equation 2.47 by setting F_i to an arbitrary value. The calculation is feed into equation 2.46, which is self-consistently refined by using the result to adjust equation 2.47, which in turn is passed back to equation 2.46. This continues until a statistical threshold is satisfied.

$$P(\xi) = \frac{\sum_{i=1}^{N_w} n_i P_i^{(b)}(\xi)}{\sum_{j=1}^{N_w} n_j e^{-\beta(w_j(\xi)-F_j)}}, \tag{2.46}$$

$$e^{-\beta F_i} = \int e^{-\beta w_i(\xi)} P(\xi) d\xi, \tag{2.47}$$

where $i = 1, 2, ..., N_w$ are the number of umbrella windows, $j = 1, 2, ..., M$ are the number of histogram bins, and $P_i^{(b)}$ is the biased probability (equation 2.44) of finding the system at $\xi = \xi_j$ in window i. In equation 2.49, k is the simulation window. When discretised, the above equations yields:

$$P(\xi_j) = \frac{\sum_{i=1}^{N_w} n_i P_i^{(b)}(\xi_j)}{\sum_{i=1}^{N_w} \frac{n_i}{\mu_i} e^{-\beta w_k(\xi_i)}} \tag{2.48}$$

$$\mu_k = \sum_{i=1}^{M} P(\xi_i) e^{-\beta w_k(\xi_i)} \tag{2.49}$$

The statistical threshold is satisfied using the following:

$$\sum_{j=1}^{N_w} \left(1 - \frac{\mu_j^{new}}{\mu_j^{old}}\right)^2 < \epsilon, \tag{2.50}$$

where the tolerance, ϵ, is typically set to some low value, say 10^{-7}. The unbiased potential of mean force (PMF) can then be obtained using:

$$G(\xi_j) = -\frac{1}{\beta} ln(P(\xi_j)) + G_0 \tag{2.51}$$

where G_0 is an arbitrary constant, typically used to normalise the PMF profile to some given free energy value.

Chapter 3

Experimental Methodologies

3.1 Suppliers of reagents and chemicals

Reagents used in this study were supplied by Fisher Scientific (UK), Sigman Aldritch (UK), Biorad, and Invitrogen (UK), unless otherwise stated.

3.2 Bacterial strains

The *E. coli* strain DH5α was used for cloning and long-term plasmid storage. The MBP deficient *E. coli* strain NT326 was used for TM protein insertion checks. All β-galactosidase assays and protein expressions checks were performed using the *E. coli* strain SU101. The cell lines are listed in Table 3.1.

Strain	Genotype	Reference
DH5α	supE44, ΔlacU169(ϕ80lacZΔM15), hsdR17, recA1, endA1, gryA96, thi-1, relA1	Sambrook and Russell (167)
NT326	F-(argF-lac)U169, rpsL150, relA1, rbsR, flbB5301, ptsF25, thi-1, deoC1, ΔmalE444, recA, srlA$^+$. Strep25 resistance.	Treptow and Shuman (168)
SU101	lexA71::Tn5(Def)sulA211 Δ(lacIPOZYA)169/FÕlacIq lacZÆM15::Tn9 CAM20, Km30 resistance.	Dmitrova *et al.,* (169)

Table 3.1: Bacterial *E. coli* strains used in this study.

3.3 Plasmids

The plasmids used throughout this study are listed in Table 3.2. The destination vector pBLM100 was used to genetically ligate a TM domain between LexA and MBP. The plasmids pBLM-GpA and pBLM-G83I were used as positive and negative controls for all experimental work.

Plasmid	Details	Reference
pBLM100	Carries ampicillin antibiotic resistance.	Schneider and Engelman (26)
pBLM-GpA	pBLM with glycophorin A transmembrane domain. Carries ampicillin antibiotic resistance.	
pBLM-G83I	pBLM-GpA with a G83I substitution. Carries ampicillin antibiotic resistance.	

Table 3.2: List of previously generated plasmids vectors used in this study.

3.4 Primers

Both controls, GpA and G83I, along with all polyleucine low complexity primer sequences (Table 3.3) were supplied by Invitrogen (UK). All alanine zippers sequences (Table 3.4,3.5) were supplied by Sigma Aldritch (UK). Nucleotides identical to the destination vector's DNA is in bold. The translated sequences from the TM domain's nucleotides are in italic.

Primer	Sequence 5' to 3'
L17	**CG** CTG CTG CTT CTG TTA CTG CTT CTG CTC CTG CTG CTT CTC CTG TTA CTG CTG **A** *L L L L L L L L L L L L L L L L*
L17(GG4)	**CG** CTG CTG CTT CTG TTA CTG GGT CTG CTC CTG GGC CTT CTC CTG TTA CTG CTG **A** *L L L L L L G L L L G L L L L L*
L17(GG4)G11I	**CG** CTG CTG CTT CTG TTA CTG GGT CTG CTC CTG ATT CTT CTC CTG TTA CTG CTG **A** *L L L L L L G L L L I L L L L L*
L17L9Q	**CG** CTG CTG CTT CTG TTA CTG CTT CTG CAG CTG CTG CTT CTC CTG TTA CTG CTG **A** *L L L L L L L L Q L L L L L L L L*

Table 3.3: DNA primers used for the preparation of *E. coli* polyleucine low complexity TM domains.

3.5 Growth and maintenance of *E. coli*

The following sections explain the preparation and cultivation of competent *E. coli* cells and the transformation of plasmid DNA.

Primer	Sequence 5' to 3'
AZ2	**CG** GCA CTG CTG GCA GCC CTG TTA GCA CTG CTT GCA GCC TTA CTG GCA CTG CTG **A** *A L L A A L L A L L A A L L A L L*
AZ2$(GG4)_L$	**CG** GCA CTG CTG GCA GCC GGT TTA GCA CTG GGC GCA GCC TTA CTG GCA CTG CTG **A** *A L L A A G L A L G A A L L A L L*
AZ2$(GG4)_L$G6I	**CG** GCA CTG CTG GCA GCC ATT TTA GCA CTG GGC GCA GCC TTA CTG GCA CTG CTG **A** *A L L A A I L A L G A A L L A L L*
AZ2$(GG4)_L$G10I	**CG** GCA CTG CTG GCA GCC GGT TTA GCA CTG ATT GCA GCC TTA CTG GCA CTG CTG **A** *A L L A A G L A L I A A L L A L L*
AZ2L6S	**CG** GCA CTG CTG GCA GCC AGC TTA GCA CTG CTT GCA GCC TTA CTG GCA CTG CTG **A** *A L L A A S L A L L A A L L A L L*
AZ2L10S	**CG** GCA CTG CTG GCA GCC CTG TTA GCA CTG AGC GCA GCC TTA CTG GCA CTG CTG **A** *A L L A A L L A L S A A L L A L L*

Table 3.4: DNA primers used for the preparation of *E. coli* AZ2 low complexity TM domains.

Primer	Sequence 5' to 3'
AZ2$(GG4)_A$	**CG** GCA CTG CTG GCA GCC CTG TTA GGT CTG CTT GCA GGC TTA CTG GCA CTG CTG **A** *A L L A A L L G L L A G L L A L L*
AZ2$(GG4)_A$G8I	**CG** GCA CTG CTG GCA GCC CTG TTA ATT CTG CTT GCA GGC TTA CTG GCA CTG CTG **A** *A L L A A L L I L L A G L L A L L*
AZ2$(GG4)_A$L6A	**CG** GCA CTG CTG GCA GCC GCA TTA GGT CTG CTT GCA GGC TTA CTG GCA CTG CTG **A** *A L L A A A L G L L A G L L A L L*
AZ2$(GG4)_A$L10A	**CG** GCA CTG CTG GCA GCC CTG TTA GGT CTG GCA GCA GGC TTA CTG GCA CTG CTG **A** *A L L A A L L G L A A G L L A L L*
AZ2$(GG4)_A$L6AG8I	**CG** GCA CTG CTG GCA GCC GCA TTA ATT CTG CTT GCA GGC TTA CTG GCA CTG CTG **A** *A L L A A A L I L L A G L L A L L*
AZ2$(GG4)_A$L10AG12I	**CG** GCA CTG CTG GCA GCC CTG TTA GGT CTG GCA GCA ATT TTA CTG GCA CTG CTG **A** *A L L A A L L G L A A I L L A L L*
AZ2$(GG4)_A$G8IL10A	**CG** GCA CTG CTG GCA GCC CTG TTA ATT CTG GCA GCA GGC TTA CTG GCA CTG CTG **A** *A L L A A L L I L A A G L L A L L*

Table 3.5: DNA primers used for the preparation of *E. coli* AZ2TF low complexity TM domains.

3.5.1 Media

Luria-Bertami (LB) medium for aerobic growth of *E. coli* liquid culture was made from 10 g/L bacto-tryptone, 5 g/L yeast extract and 10 g/L sodium chloride. Unless stated otherwise, a typical

overnight culture was made by inoculating 5 mL of media with bacteria in a 20 mL universal vial. The addition of 16 % (w/v) of agar was used for plates. Antibiotics were added to increase selectivity of cloned cells. The growth media for DH5α and NT326 were inoculated with ampicillin to a final concentration of 100 μg/mL. For GALLEX, media was inoculated with final concentrations of chloramphenicol 5 μg/mL, ampicillin 100μg/mL, and kanamycin 5μg/mL. For growth on media plates the concentration of antibiotics was kept the same.

3.5.2 Maintenance

Long-term storage of transformed DH5α stocks was achieved by adding 500 μL/mL 50 % glycerol to 500 μL/mL overnight cell culture with ampicillin. The original bacterial cell lines were grown and stored in the absence of antibiotics. Samples were frozen in liquid nitrogen and stored at -80°C.

3.5.3 Preparation of competent *E. coli* cells

Cells to be made competent were added to 2.5 mL LB media and grown in a shaker at 37°C overnight. A 1 mL aliquot of cell culture was added to 100 mL LB media and shaken in a conical flask at 37°C until OD_{600} reached between 0.4 and 0.6. The cell culture was transferred into two 50 mL vials and centrifuged at 3000 rpm for 10 min. The supernatant was discarded and each pellet was resuspended in 25 mL ice cold 40 mM $CaCl_2$. The cells were left on ice to equilibrate for 5 min. The cell solution was pelletted as described above. The supernatant was discarded and each pellet was resuspended in 2.5 mL 50 mM $CaCl_2$ plus 20% glycerol. The cells were left on ice to equilibrate for 30 min, during which time 500 μL tubes were chilled on ice. Aliquots of 100 μL cell solution were dispensed into pre-chilled eppendorfs and flash frozen in liquid nitrogen.

3.5.4 Transformation of competent *E. coli* cells

To prepare 75 - 100 μL competent *E. coli* cells were mixed with 1 - 5 μL purified DNA and left to incubate on ice for 30 min. Cells were then incubated at 45°C for 45 sec and immediately returned to ice for 2 min. The cell solution was inoculated with 400 μL fresh LB media and left to shake at 180 rpm for 1 hr at 37°C. Cells were harvested by centrifugation at 3000 rpm for 15 min and then plated out onto fresh LB agar plates.

3.6 DNA manipulation and cloning techniques

The following sections explain the extraction and purification of plasmid DNA, the conformation of modifications and inserts by colony PCR and agarose gel electrophoresis, restriction endonuclease digestion, ligation, and vector sequencing.

3.6.1 Preparation of plasmid DNA

Plasmid DNA was isolated from harvested cells using an QIAprep Mini-Prep kit (Quagen, Germany). The kit was used according to the manufacturer's instructions. Plasmid DNA was recovered in 30 μL ddH_2O and stored at -20°C.

3.6.2 Colony PCR

Ligated TM domain DNA inserts were amplified using colony PCR and then identified according to size using agarose gel electrophoresis before sequencing. Colonies were picked from fresh overnight LB agar plates and used to inoculate 50 μL ddH_2O. Then, 1 μL colony ddH_2O solution was added to 1 μL Taq buffer with KCl (without $MgCl_2$), 0.6 μL 50 mM $MgCl_2$, 0.1 μL 10 μM dNTPs, 0.1 μL 10 μM pABLM forward primer, 0.1 μL 10 μM pcc reverse primer, 0.5 μL W-1 detergent, 0.1 μL Taq polymerase (One Taq Hot Start), and 6.5 μL ddH_2O. A final volume of 10 μL of solution was transferred to a PCR tube. The PCR machine was set to:

Lid temperature	99°C	
Initial denaturation	95°C	3 min
For ×30 cycles:		
Denaturation	95°C	30 sec
Annealing	66°C	30 sec
Extension	68°C	1 min
Final Extension	68°C	
Hold	10°C	
End Loop		

PCR solutions were loaded into an agarose gel as described in section 3.6.3. An LB liquid media overnight was inoculated with the remaining 49 μL colony ddH_2O solution if DNA bands of the

correct molecular weight were identified.

3.6.3 Agarose gel electrophoresis

To prepare pBLM100 cut destination vectors, 1 % (w/v) agarose was dissolved in TAE buffer (400 mM Tris base, 20 mM acetic acid, 1mM EDTA with a pH of 8.0). For colony PCR amplified products, 2 % (w/v) agarose was dissolved in TAE buffer. A 1:20000 solution of the nucleic acid stain SafeView was added to the respective agarose solution. DNA samples were mixed with DNA loading dye according to the DNA to dye volume ratio (manufacturer dependent). For cut vector plasmids, 50 μL sample were loaded per well, and a 1 kb DNA ladder was used to identify bands according to weight. Colony PCR amplified products were loaded into 10 μL wells, and a 100 bp DNA ladder was used. Gels were run at 100 V, 150 mA for 35 min. DNA bands were visualised under a UV transilluminator and photographs of gels were taken with an Imaging System.

3.6.4 Purification of DNA by gel extraction

DNA fragments were visualised under UV light and exercised from the gel and then purified using a QIAprep Gel Extraction kit according to the manufacturer's instructions. DNA was recovered in 30μL ddH_2O.

3.6.5 Restriction endonuclease digestion of DNA

Restriction endonuclease digestion of pBLM100 DNA at the sites SacI and SpeI was performed according to the manufacturer's instructions of the restriction enzymes. Typically, Fermentas Fast Digestion required incubation at 37 °C for 5 - 10 min before recovering the DNA fragments using agarose gel electrophoresis (section 3.6.3) and gel extraction (section 3.6.4).

3.6.6 Ligation of DNA fragments

Ligation of DNA was performed using T4 DNA ligase. A ratio of 1:3, plasmid vector to DNA insert was used, and typically incubated at room temperature for 3 - 4 hrs or over night at 4 °C.

3.6.7 Sequencing of plasmid DNA

Ligated DNA was sequenced using the single plasmid read sequencing service at DNA Sequencing and Bioinformatics - GATC Biotech. Samples of 15 μL of DNA were prepared in 1.5 mL eppendorfs and 2 μL pcc forward primer was dispatched with every eight DNA samples.

3.7 GALLEX methods

GALLEX is an *in vivo* reporter system used to measure the homo-interactions of a TM domain in the *E.coli* inner membrane (26). The assay provides a semi-quantitative means of comparing the relative strengths of homo-dimerisation for our TM domains with low complexity backgrounds in a biological membrane. DNA encoding of TM domains are genetically ligated between the gene for the DNA binding domain of the *E. coli* LexA transcription factor, and the gene for maltose binding protein, MBP. The MBP region resulting chimera is positioned in the periplasmic region and acts as a membrane anchor. Expression of the chimeric protein is controlled by the induction of the lac promoter with IPTG (Figure 3.1) and as *E. coli* is incapable of metabolising IPTG the concentration remains constant.

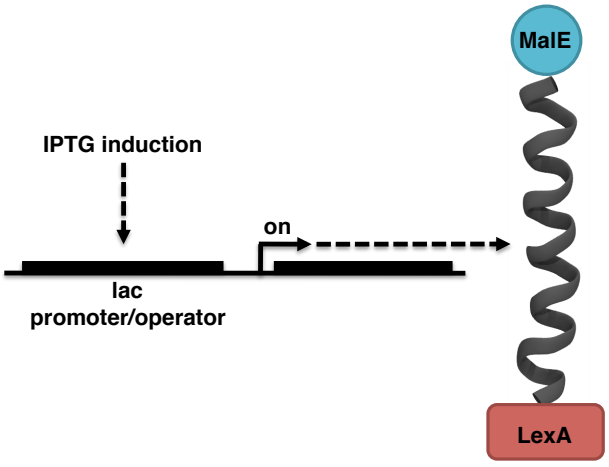

Figure 3.1: GALLEX chimeras are expressed through the binding of the IPTG transcription factor to the lac promoter.

Transmembrane helix dimerisation is measured by binding of the dimeric LexA complex to the op+/op+ promoter/operator region of the *E. coli* genome, resulting in repression of the reporter gene lacZ, and subsequent reduction in expression of β-galactosidase (Figure 3.2). This mechanism yields a system where the strength of the association between the TMs of interest is proportional to the inhibition of β-galactosidase.

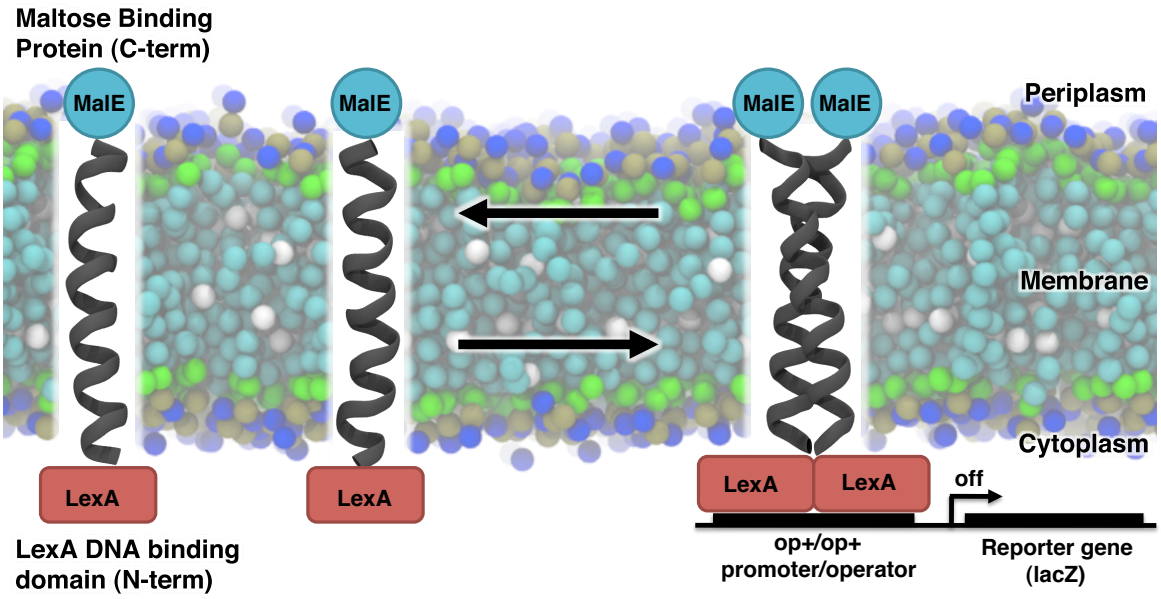

Figure 3.2: The GALLEX chimera consisting of the dimer dependant DNA binding protein LexA, the TM domain, and the periplasmic anchor MBP. The TM domain interaction enables LexA dimerisation and subsequent repression of β galactosidase by binding to the op+ promoter.

The strength of association between TM domains is visualised through a colourmetric assay. Solutions of lysed cells are inoculated with ortho-Nitrophenyl-β-galactosidase (ONPG), which hydrolyses into ortho-nitrophenol and galactose in the presence of β-galactosidase.

Controls were performed to ensure correct membrane insertion and orientation of the TM protein by transforming the ligated pBLM plasmid into competent NT326 *E. coli* cells. Some constructs were subjected to a sodium hydroxide wash to determine whether the protein was associating with the membrane, and all constructs were used in a MalE complementary assay to determine the correct orientation of the TM protein.

3.7.1 Chimera expression, insertion and orientation checks

The following sections explain the steps required to determine TM domain orientation, TM domain association with the membrane, and expression levels of chimeric protein.

3.7.1.1 SDS-PAGE

Quantification of the chimeric protein was determined using SDS-PAGE gels and western blotting. The 12% resolving layer of the polyacrylamide gels were prepared with 1.65 mL ddH_2O, 2 mL 40% acrylamide, 1.3 mL of 1.5 M Tris pH 8.8, 50 μL 10% SDS, 50 μL 10% APS, and 2 μL TEMED. The 4% stacking layer was prepared from 1.4 mL ddH_2, 0.266 mL 40% acrylamide, 0.26 mL 0.5 M Tris pH 6.8, 20 μ 10% SDS, 20 μ 10% APS and 2 μL TEMED. Samples were prepared by mixing cell solutions and lysed cell fractions with SDS sample loading buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.02% bromophenol blue, 5% β-mercaptonethanol) then boiled for at least 10 min. Running buffer was prepared with 25 mM Tris, 250 mM glycine, and 0.1% SDS. Gels were run at 125 V, 36 mA, for 50 min.

3.7.1.2 Western Blotting

A stock of 1 L TBS (Tris-Buffered Saline) was prepared with 8 g NaCl, 0. g KCl and 3 g of Tris base. The pH was adjusted to 7.4 then 0.1% Tween20 was added to make TBST. Transfer of protein from gels to the nitrocellulose membranes was performed according to the instructions for the Invitrogen and Biorad kits. The resulting blots were blocked using 2% milk TBST solution for at least 1 hr on a rocker at room temperature. Nitrocellulose membranes were then washed with TBST for 3 $\times$ 5 min. The anti-MBP mouse primary anti-body was diluted to make a 1:4000 2% milk with TBST solution and applied to the blot. The blot was left to incubate at room temperature for 1 hr on a rocker. The primary anti-body was removed with a repeat of the 3 $\times$ TBST washing. The final wash was removed and a 1:10000 dilution of an anti-mouse secondary anti-body in 2% milk and TBST solution applied to the blot and left to incubate at room temperature on a rocker. The wash was repeated as above. Protein bands were visualised by washing blots in 3 mL of BCIP/NBT reagent for 10 min. To stop the reaction the blot was washed in TBST for 10 min.

3.7.1.3 MalE complementary assay

A 5$\times$ M9 salts stock was prepared by dissolving 33.79 g Na_2HPO_4, 15 g of KH_2PO_4, 2.5 g NaCl and 5.0 g NH_4Cl in filter sterilised ddH_2O to a final volume of 500 mL. A stock of 1 M $MgSO_4$, 1 M $CaCl_2$ and 20 % w/v volume maltose were prepared in ddH_2O and filter sterilised. 78 mL ddH_2O plus 1.5 g Agar was mixed for a final agar concentration of 15 g/L before being autoclaved. Maltose

plates were prepared by adding 20 mL 5× M9 salts, 0.2 ml 1 M $MgSO_4$, 2 mL 20 % maltose, and 0.01 mL 1 M $CaCl_2$ into the agar solution. Before pouring, the stock was inoculated with IPTG for a final concentration of 0.01 M. An NT326 colony was picked from an LB ampicillin plate and streaked onto the maltose plate and then incubated for two to three days at 37°C. A segment of the plate was left blank and another segment was streaked with untransformed NT326 cells as a negative control.

3.7.1.4 Sodium hydroxide extraction assay

5 ml of LB media was inoculated with 50 μL cell culture and IPTG to a concentration of 0.01 M. The sample was incubated in a shaker until it reached an OD_{600} 0.6 to 0.8. Cell cultures were centrifuged at 3000 rpm for 10 min before being resuspended in 90 μl ddH_2O, 2.4 μL $MgCl_2$, 5 μL DNase (10 mg/ml) and 5 μL lysozyme (10 mg/mL). Samples were left to equilibrate at room temperature for 1 hr then cooled on ice. 150 μL ice cold ddH_2O was added. A 125 μL sample was extracted for analysis (whole cell fraction) and set to one side on ice. The remaining cell solution was inoculated with 125 μL ice cold 0.1 M NaOH and vortexed for 1 min. The sample was centrifuged at 14000 rpm for 15 min. Whole cell and supernatant fractions were mixed with 1 mL of 10 % TCA and left on ice for 30 min. Samples were centrifuged at 14000 rpm for 15 min and the supernatant was discarded. 1 mL of acetone was washed onto the surface of the pellet and left to incubate on ice for 5 min. A final centrifugation step at 14000 rpm for 10 min was performed. The supernatant was discarded and the pellet was allowed to air dry. The three fractions were resuspended in 80 μL 1× protein sample buffer for quantification on a western blot.

3.8 Measurement of β-galactosidase activity

The production of the β-galactosidase activity signal is described in section 3.8.1. The assay is then extended to include an IPTG concentration titration step to enable the calculation of helix-helix TM domain disassociation constants (section 3.8.2) (33).

3.8.1 β-galactosidase Assay

A sterile filtered stock of 5× Z-buffer was prepared from 300 mM Na_2HPO_4, 200 mM NaH_2PO_4, 50 mM KCl, and 5 mM $MgSO_4$. 4 mg/mL of ONPG was added to 1× Z-buffer and sonicated until

fully dissolved. Fresh 5 mL LB was inoculated with 50 μL cell culture and IPTG to a concentration of 0.01 M. Ampicillin, kanamycin and chloramphenicol were added to final concentrations of 100 μg/mL, 5 μg/mL and 5 μg/mL, respectively. Outgrowths were left to incubate in a shaker at 37°C until an OD_{600} 0.6 to 0.8. The OD was recorded and then the outgrowths were cooled on ice for the duration of the assay. 50 μL of cell culture and 50 μL growing media for the blank were aliquoted into 2 mL eppendorfs. Further preparation by adding 900 μL Z-buffer with β-mercaptoethanol was performed in a fume hood. The cells were lysed by adding 10 μL 0.1 % SDS and 2 drops of chloroform and then vortexed for 30 sec. Samples were left to equilibrate to room temperature.

Starting reactions were performed every 15 sec by inoculating one sample with 200 μL ONPG followed by brief vortexing. Samples were left to incubate for a maximum of 8 min. The procedure was repeated in the same order with 500 μL 1 M Na_2CO_3 to stop the reaction. The cells were spun down at 14000 rpm for 10 min before transferring the supernatant to cuvettes. An OD_{420} for pigmentation and OD_{550} for cell debris were recorded. The β-galactosidase miller units were calculated using equation 3.1:

$$\frac{1000(OD_{420} - 1.75 \times OD_{550})}{t \times v \times OD_{600}}, \tag{3.1}$$

where t is time in minutes and v is volume in micro litre.

3.8.2 β-galactosidase Free Energy Assay

Z-buffer, ONPG, and media growth stocks were prepared according to section 3.8.1. The apparent free energy assay was performed on a single TM chimeric protein construct at a time. Chimeric plasmid DNA was transformed as described in section 3.5.4 and then plated out onto LB agar. Three LB media overnights of SU101 colonies were prepared as described in section 3.5.1. Using the three overnights, ten sets of three independent outgrowths were prepared in 5 mL of LB media with the necessary antibiotics. To build the IPTG concentration curve, each set of three independent samples were inoculated with IPTG to a final concentration of 1 mM, 0.5 mM, 0.1 mM, 0.05 mM, 0.01 mM, 0.005 mM, 0.001 mM, 0.0005 mM, 0.0001 mM, or 500 μL plain dH_2O. Measurement of β-galactosidase activity was performed as outlined in section 3.8.1.

3.9 Specialist software

1. Serial Cloner: design of TM domain primers and manipulation of pBML plasmid destination vector.
2. 4Peaks: Analysis of sequencing data.
3. ImageJ: Analysis of expression check western blots (170).
4. ΔGpredictor: Analysis of free energy of TM peptide insertion .
5. OriginPro: GALLEX analysis and data visualisation.

Chapter 4

Lipid-Mediated Dimerisation of the Transmembrane Helical Peptide Neu: Insights from Molecular Dynamics Simulations

4.1 Introduction

In this chapter we explore two distinct areas of helix-helix self-association from a top-down perspective: how the bilayer membrane composition and the effects of a single substitution can affect the structure and behaviour of a TM peptide.

How much of an affect does the membrane-mimetic have on the structure of a TM peptide? As a poignant example, a number of research groups, each reaching a different conclusion, have explored the structure of TM helix 2 of α_1 glycine receptor. The target peptide was studied in: TFE isotropic solvent (171, 172); SDS detergent micelles and DPC (dodecylphosphocholine) detergent micelles (173); and DMPC lipid bilayers (173). Many simulations in the study by Johnston *et al.,* (6), verified that there is structural variation of the same helical peptide when inserted into these different membrane-mimetics.

A further poignant example utilises MD to study the free energy of self-association of the receptor tyrosine kinases human ErB family (96) in different bilayer membrane compositions. Each peptide was separated by 20 Å to generate a reaction coordinate over which umbrella sampling was performed to yield PMF profiles. The variation in free energy between membrane-mimetic simulations indicated that the protein-protein interactions at short-range were drastically different in membranes of different compositions. Unfortunately, the simulations failed to probe the effects of lipid-mediated long-range interactions between TM peptides.

In order to investigate the short-range and long-range effects critical for driving membrane protein oligomerisation in a membrane bilayer, without losing any detailed information on lipid-lipid interactions, we have used atomistic MD simulations and free energy calculations. The helical TM domain of the rat Neu protein, an extensively-studied receptor tyrosine kinase, was used as a model TM protein in this work and was studied in a single component POPC (full name 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer. The oncogenic form of Neu, containing a transmembrane $V_{664}E$ substitution that results in permanent activation of its catalytic kinase domain (174, 175, 176, 177), was also investigated.

Recent data obtained from the inner membrane of Escherichia coli using the TOXCAT assay (25) suggests that the oligomerisation of Neu is stabilised by two highly conserved motifs located on opposite sides of the helix (178). The first *inter*helical motif, denoted Neu_{AG}, consists of the sequence $A_{661}xxxG_{665}$, which is known as the Sternberg-Gullick motif (74). The second *inter*helical motif, denoted Neu_{IV}, consists of the sequence of $I_{659}xxxV_{663}$. It has been suggested that the helix-helix orientation ($I_{659}xxxV_{663}$ or $A_{661}xxxG_{665}$ at the helix-helix interface) acts as a *molecular switch* (179, 180, 181, 182), causing a conformational change in a preformed, inactive dimer that leads to activation of the kinase domain in the cytoplasmic domain (183, 184).

Our results demonstrate that the length of the Neu TM domain gives rise to positive hydrophobic mismatch and causes stretching and ordering of the lipids surrounding the protein to compensate for this. Simulations also reveal that the single amino acid substitution drastically affects free energy values and *inter*helical distance minima. Most significantly, our free energy calculations demonstrate changes in free energy at *inter*helical distances well beyond those for direct protein-protein interactions (> 6 nm) and suggest a significant long-range entropic contribution to membrane protein oligomerisation that is transmitted via the lipid bilayer.

In addition to the effect of lipid mediated protein-protein interactions, we explore how the proto-oncogenic and oncogenic forms of Neu affect the thermodynamic characteristics of the two oligomeric states.

4.2 Methods

4.2.1 Forcefield parameters

The peptides were modelled using the united atom Gromos53a6 force field (185) which, has been shown to give a reasonable estimation of peptide backbone ϕ- and ψ-dihedral angle probability distributions compared with other force fields (186). Water molecules were represented using the Simple Point Charge (SPC) model (187, 188). Compatible chloride counterions were used to neutralise the five positive arganine side chains on each peptide. POPC lipids were modelled using a Gromos53a6 compatible lipid force field (189).

4.2.2 Simulation parameters

Simulations were performed using Gromacs 4.5.5 (190). All bonds were constrained using the LINCS algorithm (155) and thus a time step of 0.002 ps was employed. The temperature of the system was maintained at 323 K using the Nosé-Hoover thermostat (150). This temperature was chosen to ensure that the bilayer was in the liquid crystalline phase and also to enable comparison with our previous simulations of Neu in a DPPC bilayer (46). The Parrinello-Rahman pressure-coupling scheme (153) was used to maintain the pressure at 1 bar. The pressure-coupling was applied in a semi-isotropic manner, whereby the z-dimension of the simulation cell (direction of the bilayer normal) was scaled independently from the xy-dimensions (the plane of the bilayer). Long-range electrostatic interactions were treated using the particle mesh Ewald approach (127) with a short-range cut off of 1.4 nm. Lennard-Jones interactions were cut off at 1.4 nm and terminated with a simple cut off.

4.2.3 Generation of initial configurations of the dimer

The wild-type Neu TM domain has the sequence (R_{652}ASPVTFIIATVV$_{664}$GVLLFLILVVVVGILIKRRR) and the oncogenic form, denoted Neu*, contains the V_{664}E point mutation. For clarity, proto-

oncogenic Neu with an initial starting configuration where $I_{659}xxxV_{663}$ or $A_{661}xxxG_{665}$ represents the primary binding interface will be referred to as Neu_{IV} or Neu_{AG}, respectively. This is also the case for Neu*. To generate starting structures for MD simulations, the Neu and Neu* sequences were subjected to a global conformational search in a low dielectric environment using the CHI algorithm (191). The centre-of-mass *inter*helical distance was set to 1.15 nm, in agreement to the optimum distance found by previous conformational searches of Neu (192). Right and left-handed crossing angles, as defined by the angle between the principle axes of each helix, were set to -25° and 25°, respectively. The helices were rotated about their principle axes from 0° to 360° in increments of 10°. Each rotational step comprised four trial MD simulations of 5000 time steps; for each trial a different set of atomic velocities were assigned at random from a Maxwell-Boltzmann distribution. From the search results, distinct structures were collected using the CHI cluster calculation tool. The cut off for the root mean squared difference between candidate structures was 0.1 nm. Up to eight structures within this cut off constitute a complete cluster. The atomistic positions of each member of the cluster are then averaged to yield a set of possible starting configurations for further MD simulations. Homo-dimers were selected according to the nearest like for like amino acid potential energy profile and capped with methyl acetamide and methyl amid at the N- and C-terminal, respectively. Lysine and arganine were left charged due to their proximity with the lipid head group and bulk water. The glutamic acid in Neu* was protonated as per experimental data on the protonation state of glutamic acid in the membrane hydrocarbon tail region (193).

4.2.4 Generation of Initial Bilayer

A POPC bilayer was selected, as its structural properties have been parameterised well (194), and it has been used in other computational studies (46) and biophysical experiments (46) on Neu.The POPC bilayer thickness (37.9 Å) is similar to the thickness of the E. coli inner membrane (195), enabling a comparison to *in vivo* studies (46). A POPC molecule coordinate file and force field parameters were obtained from the lipid force field repository LipidBook (196). The POPC bilayer was created by aligning a single POPC molecule along the *z*-direction and then replicating the molecule in the *x*- and *y*-directions to form a 200 lipid monolayer. The monolayer was then replicated and rotated to yield a bilayer structure. The bilayer was solvated with 23765 water molecules and 10 chloride counterions (to neutralise the charge as the peptides). To ensure that the

peptides would not interact with their periodic images during the PMF calculations the bilayer was chosen to approximately 8 nm and 15 nm in the *x*- and *y*-dimensions, respectively. The bilayer was then equilibrated as per the method outlined in section 4.2.6.

4.2.5 Insertion of the dimer into the bilayer

The prepared homo-dimers were inserted into an equilibrated POPC bilayer (see section 4.2.6) using the InflateGro tool (197). Firstly, all water was temporarily removed from the system and the dimer of interest was inserted into the centre of the bilayer. The tool then translates every lipid outward, resulting in a very porous bilayer. An energy minimisation step is performed using the steepest descent algorithm with a maximum of 5000 steps, an energy threshold of 1000 kJ mol^{-1} nm^{-1} and a step size of 0.002 ps. The bilayer is then compressed by applying a scaling factor of 0.95 to the current *x* and *y* coordinates of the lipids. The closer the lipids are to their initial starting configuration, the smaller the steps become and another energy minimisation step is performed. This procedure repeats until the area-per-lipid through compression of the bilayer approximates the original area-per-lipid value prior to peptide insertion. The water is returned, counter ions are added to neutralise the net charge in the system, and the bilayer is re-equilibrated along with the peptides as described below.

4.2.6 Equilibration procedure

To equilibrate the bilayer prior to peptide insertion, a 1000 kJ mol^{-1} position restraint was applied to the phosphorus atom of each lipid head group. A 500 ps simulation was performed under constant volume. Restraints were removed and the system was allowed to run for 30 ns under constant pressure. The bilayer thickness and area per lipid were monitored during the bilayer equilibration simulation to ensure the bilayer had equilibrated to observed experimental values. After insertion of the dimer, the system was re-equilibrated. First, the system was subjected to an energy minimisation step using the steepest descent algorithm with an energy threshold of 1000 kJ mol^{-1} nm^{-1} and a step size of 0.002 ps. Position restraints of 1000 kJ mol^{-1} were then applied to peptide heavy atoms and the system was simulated for 10 ns under constant *NPT* conditions. This was followed by further 10 ns simulations each with position restraints of 100 kJ mol^{-1} and 10 kJ mol^{-1} on the heavy atoms of the peptides and a final 10 ns simulation without restraints. The orientation of each

dimer was checked to ensure that the respective packing motif (A_{661}xxxG_{665} or I_{659}xxxV_{663}) had remained at the peptide-peptide interface after the equilibration stages. Simulations were checked to ensure that the bilayer area per lipid and bilayer thickness had converged and approximated observed biological values.

4.2.7 Umbrella sampling and free energy calculations

A reaction coordinate was defined by displacing the equilibrated helices by 1 Å increments up to 75 Å per free energy profile in the *y*-dimension, yielding a total of 300 umbrella windows across all four systems of Neu. This resembles a reversible *together-apart* process. Individual intervals were equilibrated for 5 ns with a 1000 kJ mol^{-1} position restraint assigned to the heavy atoms of the peptides. The total equilibration time over all four Neu systems was 1500 ns. Position restraints were removed and an umbrella potential of 1000 kJ mol^{-1} was applied to the *xy*-lateral distance between the centre-of-mass of each peptide. Each window was then simulated for 40 ns, resulting in 2800 ns of umbrella sampling simulation per Neu system and a total time of 11200 ns. The WHAM, (as implemented in Gromacs 4.5.5) (198), was used to construct free energy profiles from the final 30 ns of trajectory. Bayesian bootstrapping was used to calculate statistical uncertainty over 200 bootstrap iterations and the average free energy was taken as the final result. The free energy profile was normalised to 0 kJ mol^{-1} where there was no longer a change in free energy (i.e., when the helices were far apart and non-interacting). This began after approximately 7.0 nm along the reaction coordinate for all four systems. To account for variation in the start of the free energy plateau between systems 7.5 nm was selected.

4.3 Results

Neu and Neu* homo-dimer α-helical structures were generated using the CHI software as described in section 4.2.3 and four conformations were selected for the subsequent MD simulations, where the A_{661}xxxG_{665} or I_{659}xxxV_{663} interface represents the primary binding interface. The structural properties (area per lipid, bilayer thickness and lipid tail order parameters) were analysed for both the pure POPC bilayer without the peptides and the POPC bilayer with the peptides at different inter-helical distances.

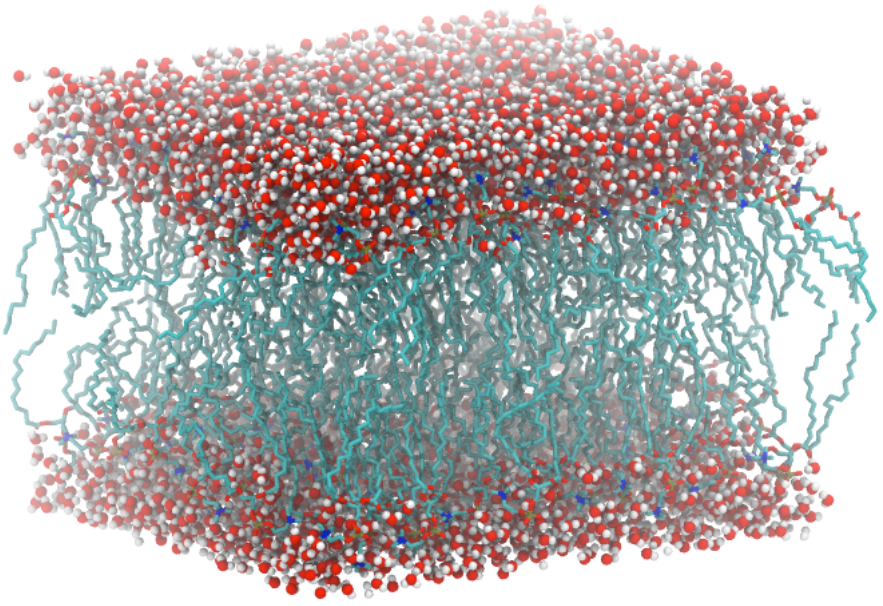

Figure 4.1: Final frame of a 35 ns simulation of POPC fully solvated in simple point charge water molecules. Only water molecules within 1.5 nm of the average lipid head group position are shown for clarity.

4.3.1 Bulk structural properties of a POPC lipid bilayer

A POPC model membrane was used as an environment to calculate the free energy of association between both sets of Neu TM domains. As a benchmark, a bulk POPC model bilayer was simulated at 323 K for 35 ns using the NPT ensemble. A final frame snapshot is presented in Figure 4.1. The bilayer thickness and area per lipid were calculated over the last 5 ns as described below.

4.3.1.1 Bilayer thickness

The simulated bilayer thickness of a bulk POPC model lipid bilayer was calculated using the GridMAT-MD software package (199). A 25×24 grid of evenly spaced points was applied parallel to both bilayer surfaces. The fluctuation to the simulation cell volume was small enough to ensure the constant arrangement of the grid (see Figure 4.3) would approximate the dimension of the simulation cell *xy*-plane. The *inter*leaflet distance between lipid phosphate groups within each grid-square was calculated per recorded frame. The reported average bilayer thickness within each grid-square was averaged over 5 ns of recorded trajectory. A range of bilayer thickness values between 3.5 nm to 4.2 nm were calculated (see Figure 4.2), yet the variation in thickness was shown to be statistically insignificant due to the standard deviation (Figure 4.2). This is indicative

of the fluidic nature of a liquid-crystalline lipid bilayer. The average bilayer thickness at each grid-square calculated over the analysed trajectory was 38.9 Å with a standard deviation of 2.9 Å. The simulated bilayer thickness compares well with the bilayer thickness of 37.9 Å at 323 K according to small-angle neutron and X-ray scattering data (194).

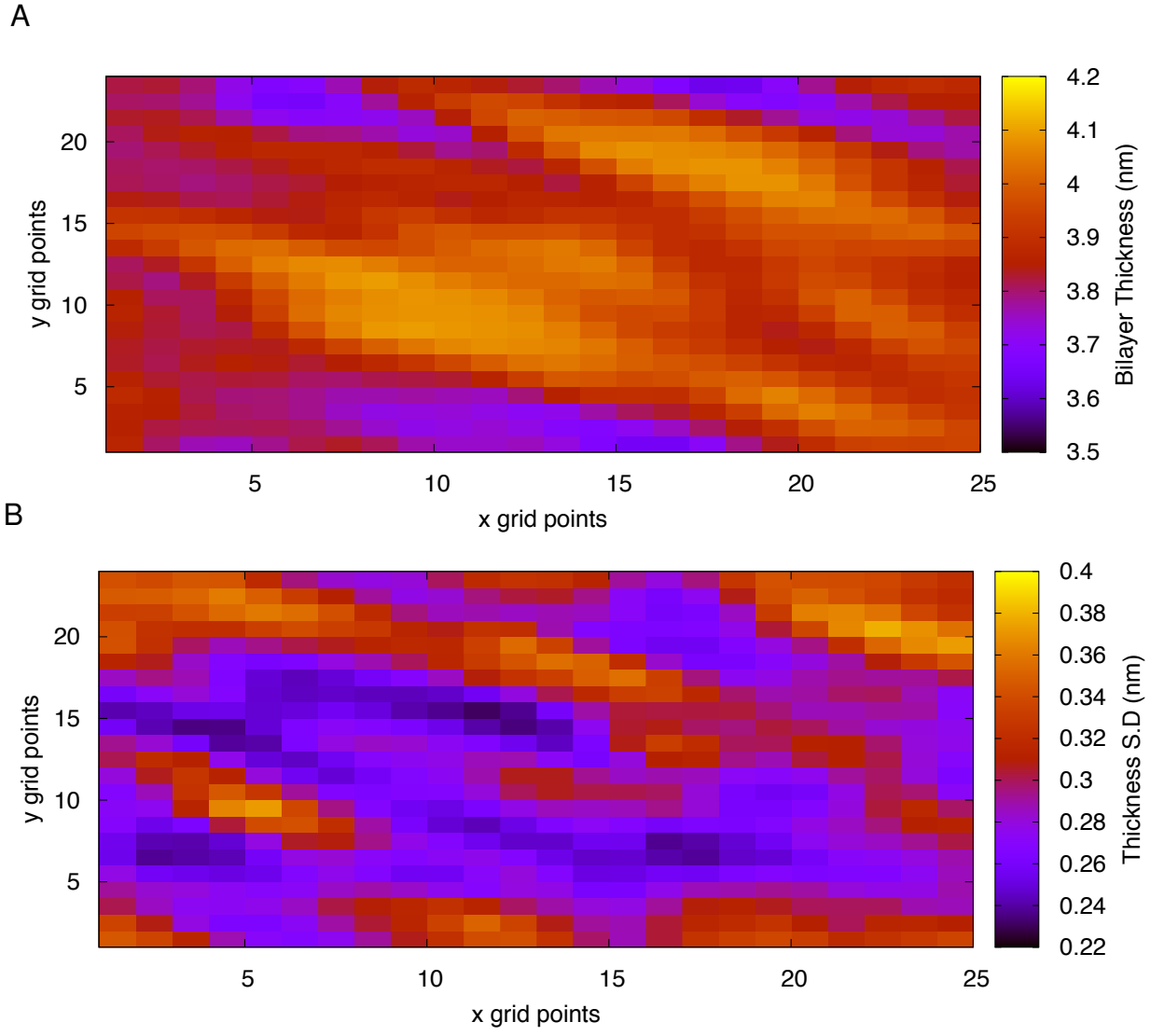

Figure 4.2: (A) POPC average bilayer thickness along with (B) the standard deviation (S.D), at each grid-square.

4.3.1.2 Area-per-lipid

The average area-per-lipid of the bulk POPC lipid bilayer was calculated using the GridMAD-MD software package (199). As with the bilayer thickness calculations, a grid of evenly spaced points was generated parallel to both leaflets, using the lipid phosphate group as a reference. The average

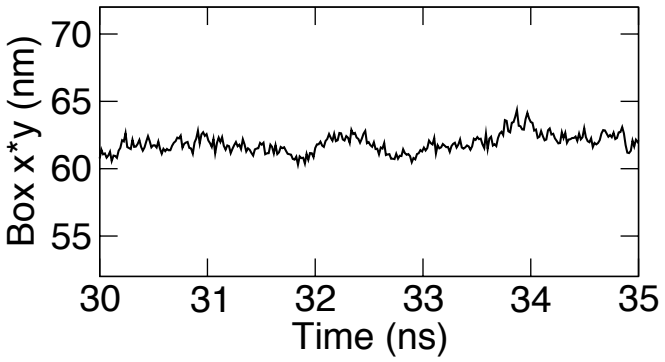

Figure 4.3: The simulation box *xy*-area, parallel to the bilayer. Standard deviation was calculated for the error, however, it was significantly small to ignore.

area-per-lipid was approximately 60.673 $Å^2$ with a standard deviation of 0.001 $Å^2$. This compares well with the experimental value of 67.4$Å^2$ ± 1.3$Å^2$ according to small-angle neutron and X-ray scattering. As presented in Figure 4.3, the converged box area suggests that the area-per-lipid only fluctuate within the confines of a small area, indicative of an equilibrated lipid bilayer.

4.3.2 Free energy profiles as a function of peptide separation

The free energy profiles, showing the relative change in the free energy of the four systems as a function of the distance between the centres of mass of the helices, are shown in Figure 4.4 The quality of umbrella histogram distribution and overlap can be seen in Figure 4.5. Further more, each PMF profile was checked for PMF convergence by performing a series of WHAM calculations over additive 4 ns increments, i.e., from 0 ns – 4 ns, 0 ns – 8 ns,..., to 0 ns – 40 ns. In all four PMF profiles, the free energy plateaus between 6.5 nm and 7.5 nm, thus each profile was normalised to ΔG = 0 kJ mol^{-1} at 7.5 nm. This *inter*helical distance of 6.5 nm is well beyond the cut off distance for van der Waals interactions between the peptides and the electrostatic interactions are also likely to be negligible. Thus, the peptides are considered to be in a monomeric state.

Starting from the monomeric state, as the peptides come into closer proximity - between 6.5 nm and 2 nm - there is an overall decrease in the free energy. This indicates that over this distance range, in all cases, there is a thermodynamic force driving the system towards the dimer state. The free energy change associated with bringing the proto-oncogenic interfaces Neu_{IV} and Neu_{AG} to their global free energy minimum is -108± kJ mol^{-1} and -97±10 kJ mol^{-1} respectively. In contrast, the free energy change associated with the oncogenic interfaces Neu^*_{IV} and Neu^*_{AG} is -55±9 kJ mol^{-1} and -58±8 kJ mol^{-1} respectively. In addition, Neu^*_{IV} features a free energy plateau from 4 nm to 6 nm at an approximate ΔG of -20 kJ mol^{-1}, and Neu_{AG} features a free energy plateau from 1.8 nm to 3.8 nm with an approximate ΔG of -40 kJ mol^{-1}.

Individual peptide and dimer tilt angles were recorded according to Appendix A over the last 20 ns of three umbrella windows per system: as a dimer (at the free energy minimum); as the peptides approach (4 nm along the reaction coordinates); and as a monomer (7 nm along the reaction coordinate). Even though a peptide tilt angle from the monomeric and approaching configurations were shown to overlap (Figure 4.6 (A)), the Neu_{IV} tilt angles were found to adopt an approach < dimer < monomer average tilt angle order. With tilt angles between 32° and 60°, the three umbrella windows of Neu_{IV} were shown to be more similar to one another than the tilt angles of Neu_{AG} which were between 10° and 60°. Although the Neu_{AG} title angles (Figure 4.6 (B)) were found to adopt a monomer < dimer < approach average tilt angle order, one peptide tilt angle from the monomeric and approaching umbrella windows were indistinguishable from one another and the

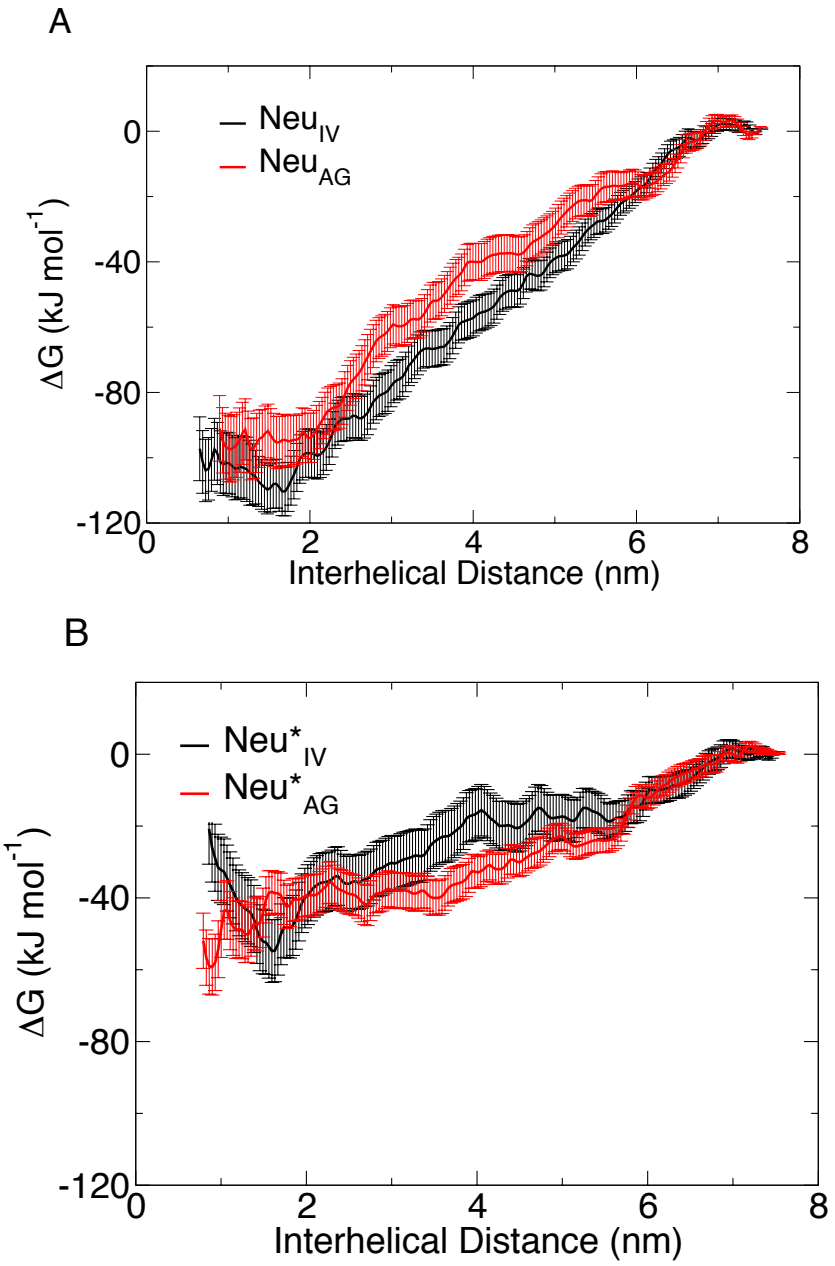

Figure 4.4: Relative free energy of association between pairs of (A) Neu_{AG} and Neu_{IV} peptides and (B) Neu^*_{AG} and Neu^*_{IV} peptides, as a function of the distance between the centre-of-mass of the two helices. The error bars were calculated using Bayesian bootstrap analysis.

dimeric tilt angle. The three sets of tilt angles in the Neu^*_{IV} umbrella windows (Figure 4.6 (C)) returned the order approach < monomer < dimer, although the last 5 ns displayed a visible overlap between the tilt angles of one of the approaching peptides with both monomeric peptides. The umbrella windows shared a similar range of tilt angles to the proto-oncogenic form, Neu_{IV}. Finally, with the exception of the overlap between the tilt angle of an approaching peptide with the tilt angles of both monomeric peptides, the average tilt order for the Neu^*_{IV} umbrella windows was

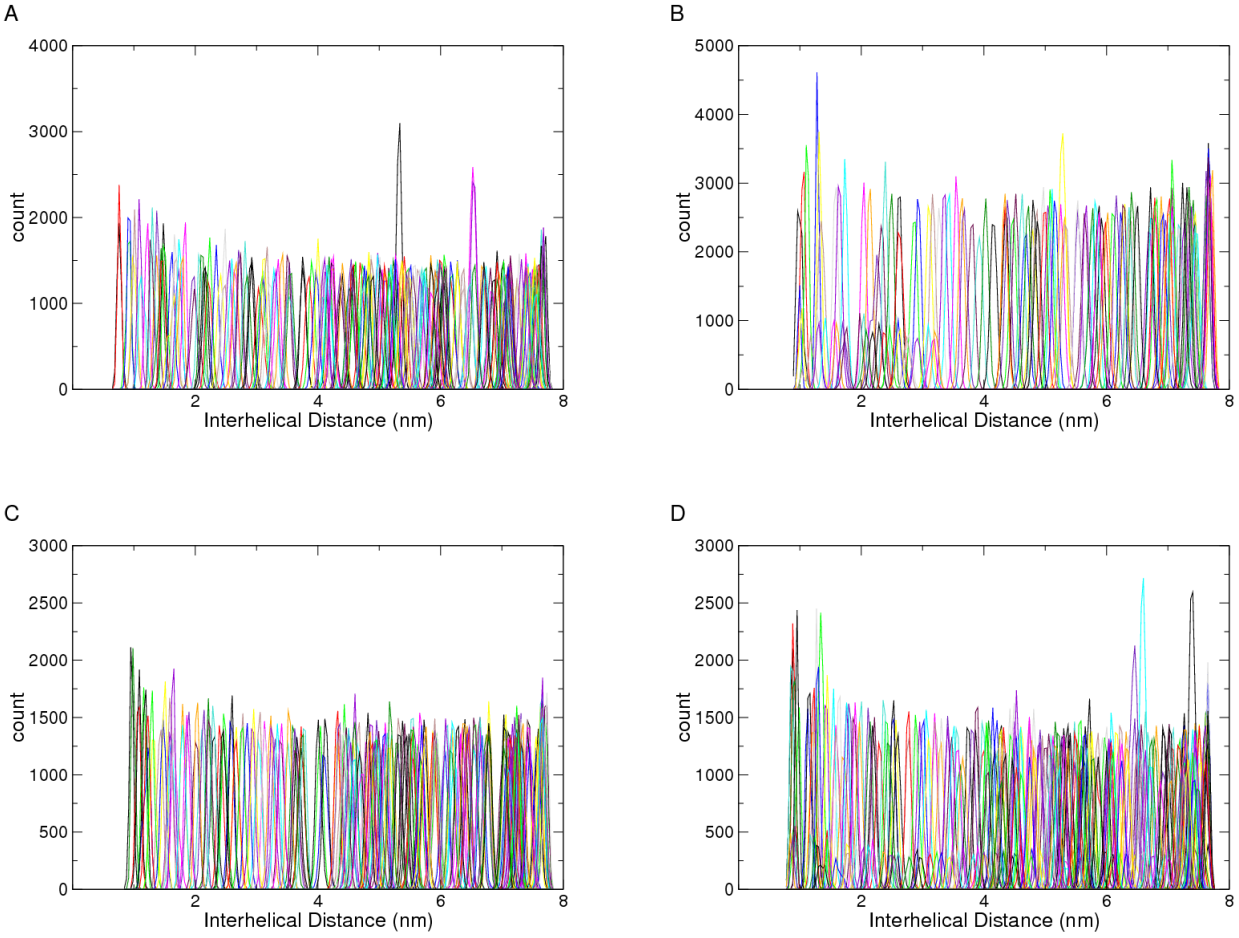

Figure 4.5: Umbrella sampling histograms along an *inter*helical distance reaction coordinate for the Neu dimers: (A) Neu_{IV}, (B) Neu_{AG}, (C) Neu^*_{IV}, and (D) Neu^*_{AG}. **Note**: Although there may appear to be an obvious gap to the naked eye between histograms approximately 4 nm in to the reaction coordinate of (C), there are approximately 200 data points of over overlap between histograms.

dimer < monomer < approach (Figure 4.6 (D)). With tilt angles between 18° and 60°, the three umbrella windows of Neu^*_{AG}, were similar the range of tilt angles from the proto-oncogenic form, Neu_{AG}.

Over the range 6.5 nm - 2 nm the peptides do not make contact with each other and all atom pairs are beyond the cut off distance for short-range interactions. Thus it was surprising to see the attraction between the peptides extend over such a long-range. One possibility is that the lipids themselves play a role in dimerization. To explore this further we investigated the molecular arrangement of the lipids as a function of their distance to the helices.

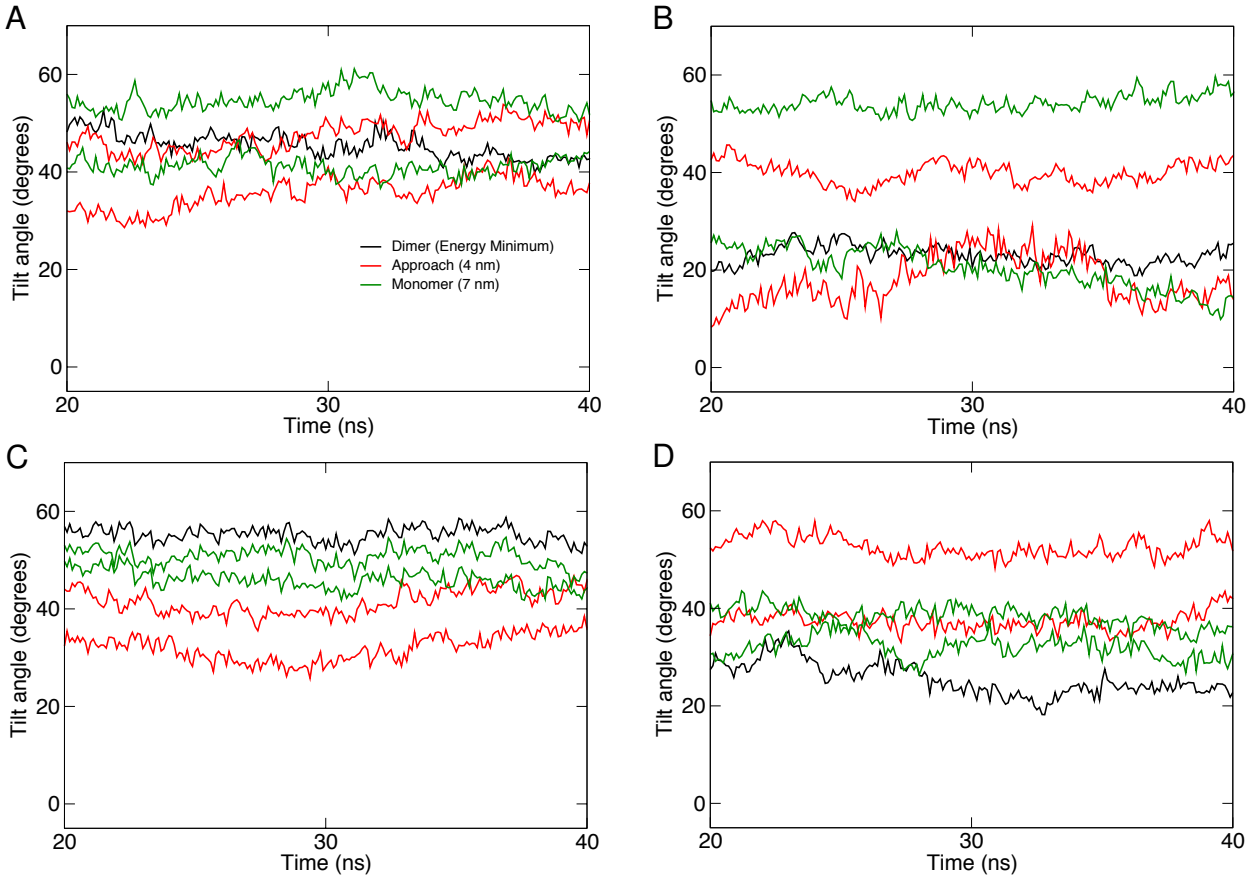

Figure 4.6: Dimer and individual peptide tilt angles of (A) Neu_{IV}, (B) Neu_{AG}, (C) Neu^*_{IV}, and (D) Neu^*_{AG}

4.3.3 Bilayer thickness

Average bilayer thickness calculations were performed for all four systems (Neu_{IV}, Neu^*_{IV}, Neu_{AG} and Neu^*_{AG}) at three selected *inter*helical distances, which represent distinct regions on the free energy profile (Figure 4.7 and Figure 4.8). These were: (i) where the *inter*helical distance was 7 nm and the peptides were in the monomeric state; (ii) where the *inter*helical distance was 4 nm and the free energy is intermediate between the dimeric and monomeric states; and (iii) where the *inter*helical distance corresponded to the global free energy minimum position. Corresponding snapshots of the peptides at each distance are also shown. It can be seen that the presence of a Neu peptide in the bilayer results in a change of thickness to the bilayer from a bulk reference of 3.8 nm to between 4.4 nm and 4.9 nm. The peak in the increase in thickness correlates well with the position of both peptides. For peptides in the monomeric state, the perturbation extends out to 2.5 nm from the peptide's centre of geometry. Beyond this distance, the bilayer adopts a bulk lipid arrangement with a thickness comparable to approximately 3.8 nm, as reported in section 4.3.1.1, suggesting that each peptide is outside the influence of the perturbed lipids of the other peptide. When the peptides are at an intermediate *inter*helical distance of 4 nm, the increase in bilayer thickness extends across the whole bilayer. At the free energy minimum position, the perturbation has an effective radius of approximately 3.5 nm from the dimer centre.

4.3.4 Lipid tail order parameters

An increase in the thickness of a bilayer often goes hand-in-hand with an increase in the order of the lipid tails (200). Essentially, tightly-packed lipid tails show an increase in lipid tail order and as a result the thickness of the bilayer increases. To investigate whether the aforementioned Neu and Neu* peptides also affect the lipid tail order, we divided each bilayer into five segments of approximately 3 nm in length along the *y*-dimension of the simulation cell. The average lipid tail order parameter of lipids in each segment was then calculated. The results for the Neu_{IV}, Neu^*_{IV}, Neu_{AG}, and Neu^*_{AG} are presented in Figures 4.9, 4.11, 4.10, and 4.12. When the peptides are in the monomeric state, with an *inter*helical distance of 7 nm, they are located in the first (0 nm – 3 nm) and fifth (12 nm – 15 nm) segments. When the peptides are 4 nm apart, they are located in the second (3 nm – 6 nm) and fourth (9 nm – 12 nm) segments. Finally, the peptides taken from the lowest free energy trajectory are located in the third segment (6 nm – 9 nm).

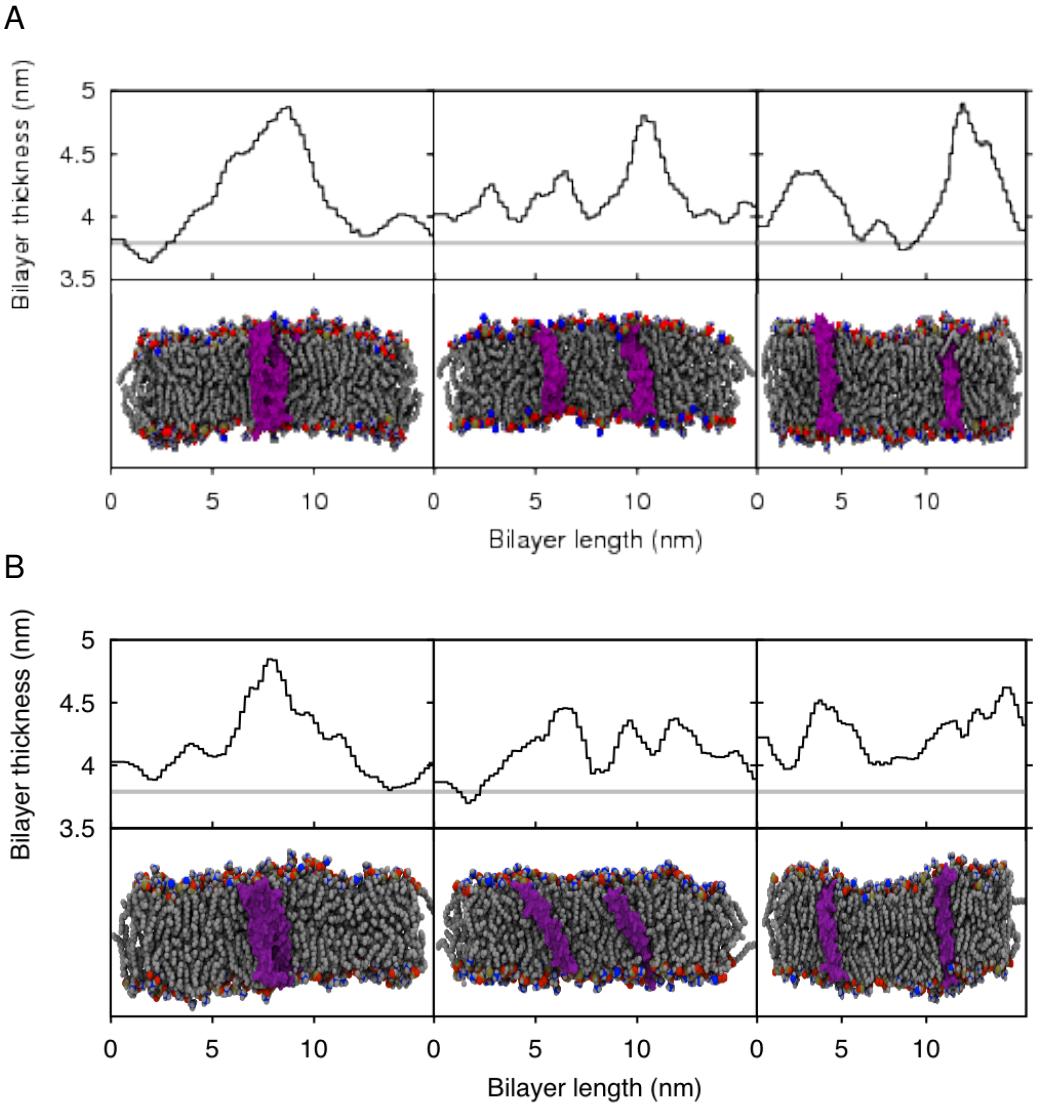

Figure 4.7: The averaged POPC bilayer thickness from three (A) Neu_{AG} and, (B) Neu^*_{AG}, simulations. From left to right: self-associated peptides at the free energy minima, *inter*helical distance of 4 nm, and an *inter*helical distance of 7 nm (monomeric). The grey line indicates an average simulation POPC bulk thickness of 3.8 nm. Lipids are in grey space filling spheres, along with head group atoms: oxygen in red, nitrogen in blue, and phosphorus in tan. Peptides are in purple surf. Water was removed for clarity.

When the peptides are in the monomeric state (blue lines), with an *inter*helical distance of 7 nm, they are located in the first (0 nm – 3 nm) and fifth (12 nm –15 nm) segments and there is an increase in the order of the lipids in these segments compared to a bulk POPC (black). When the peptides are 4 nm apart (green lines), they are located in the second (3 nm – 6 nm) and fourth (9 nm – 12 nm) segments and there is an increase in the lipid tail order that spans from the second to the fifth segment. Again this demonstrates a close correlation between the bilayer thickness and the lipid order, although there are a three noticeable exceptions, where the lipids in either segment become more ordered whilst the system maintains a pair of monomeric peptides (Figure 4.9 (A) and

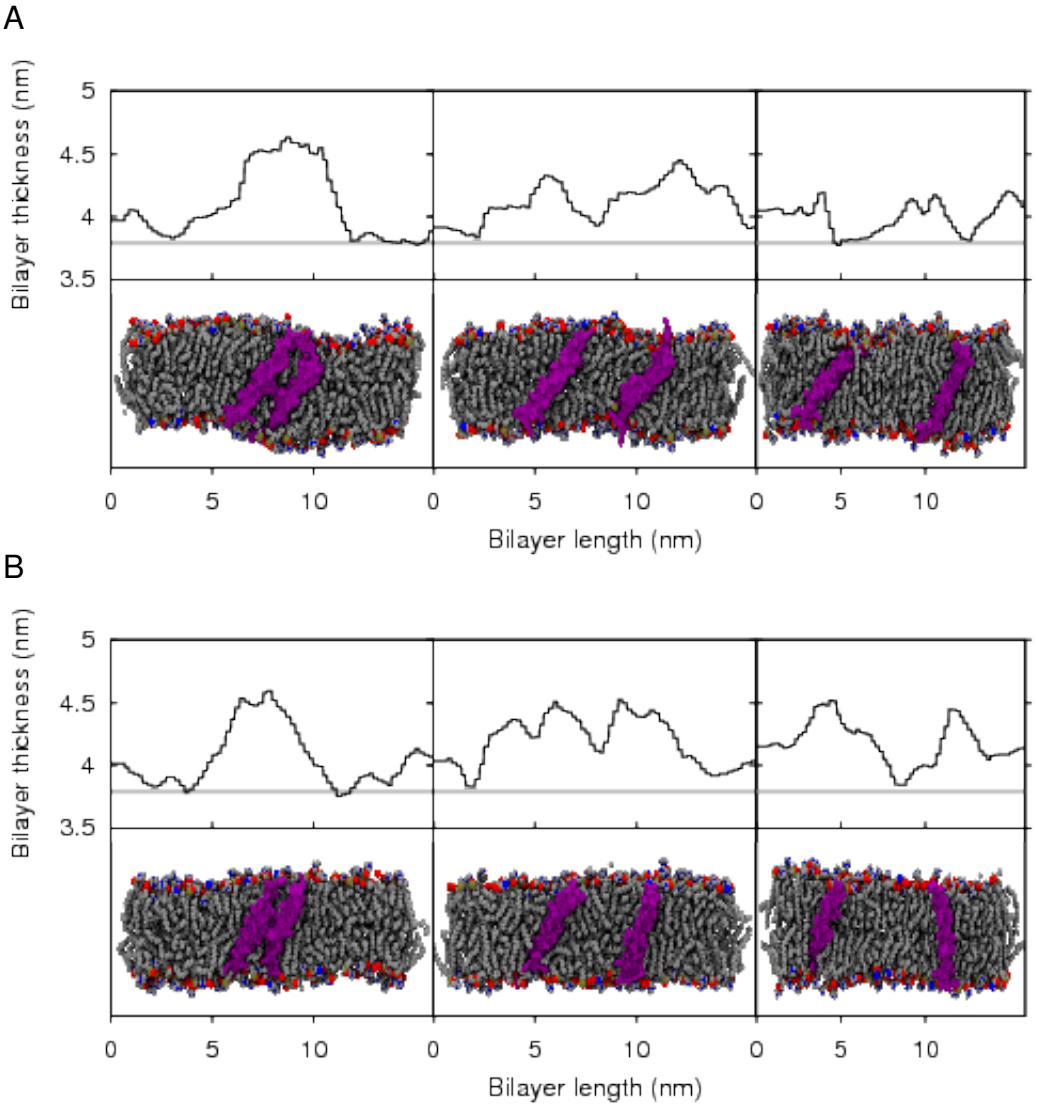

Figure 4.8: The averaged POPC bilayer thickness from three (A) Neu_{IV} and, (B) Neu^*_{IV}, simulations. From left to right: self-associated peptides at the free energy minima, *inter*helical distance of 4 nm, and an *inter*helical distance of 7 nm (monomeric). The grey line indicates an average simulation POPC bilayer bulk thickness of 3.8 nm. Lipids are in grey space filling spheres, along with head group atoms: oxygen in red, nitrogen in blue, and phosphorus in tan. Peptides are in purple surface representation. Water was removed for clarity.

(B) in segment B, Figure 4.10 (A) and (B) in segment D, and Figure 4.12 (A) and (B) in segment B). Finally, the peptides in the dimer state (red lines) are both located in the third (6 nm – 9 nm) segment. The lipid tail order is elevated in this segment and drops back to the bulk value with increasing distance from the peptides. This matches well with the decrease in bilayer thickness, which eventually reaches the bulk bilayer thickness, as indicated in section 4.3.1.1.

The most likely origin of the increase in lipid order and bilayer thickness is the hydrophobic mismatch between the peptide and bilayer. Measuring the longest span of unbroken hydrophobic residues, F_{658} to I_{680}, the average hydrophobic length of both helices as a function of time was

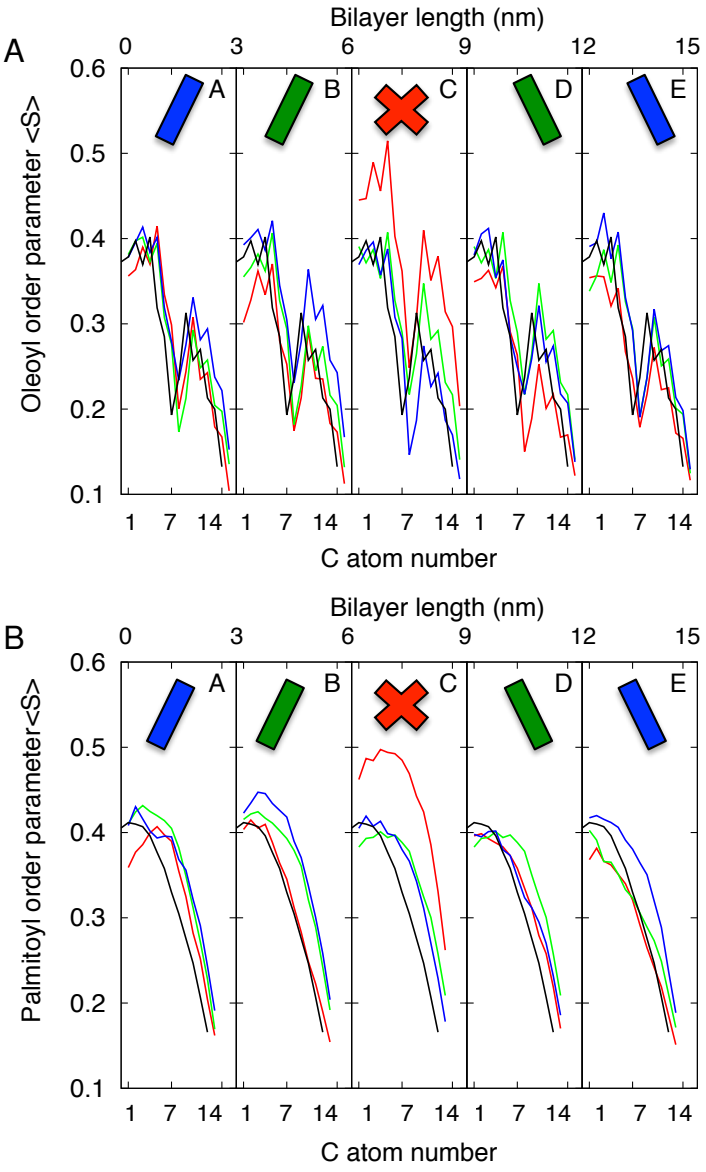

Figure 4.9: Average lipid order of (A) oleoyl tail, and (B) palmitoyl tail, from five regions of the bilayer (0 nm - 3 nm, 3 nm - 6 nm, 6 nm - 9 nm, and 12 nm - 15 nm as partitions A through to E, respectively) during simulations of Neu_{IV} as a monomer (blue), at an intermediate distance of approximately 4 nm between helices (green), and at the free energy global minimum (red). The position of each peptide is indicated by the corresponding marker. Lipid tail order calculations from a POPC bulk bilayer is shown in black. Standard deviation was calculated but the values were significantly small enough to exclude error bars.

approximately 3.1 nm to 3.4 nm at the global free energy minimum (see Figure 4.13). Relative to the thickness of the POPC hydrophobic tails (2.81 nm ± 0.6 nm (194)), the significant difference would require an adjustment to the peptide tilt angle and a change in the bilayer thickness.

In the next section we describe how close range interactions between Neu peptides was shown to be statistical insignificant in free energy according to Bayesian bootstrap analysis. Yet, by introducing the $V_{664}E$ substitution the self-association of peptides is in favour of an $A_{661}xxxG_{665}$ orientation. It

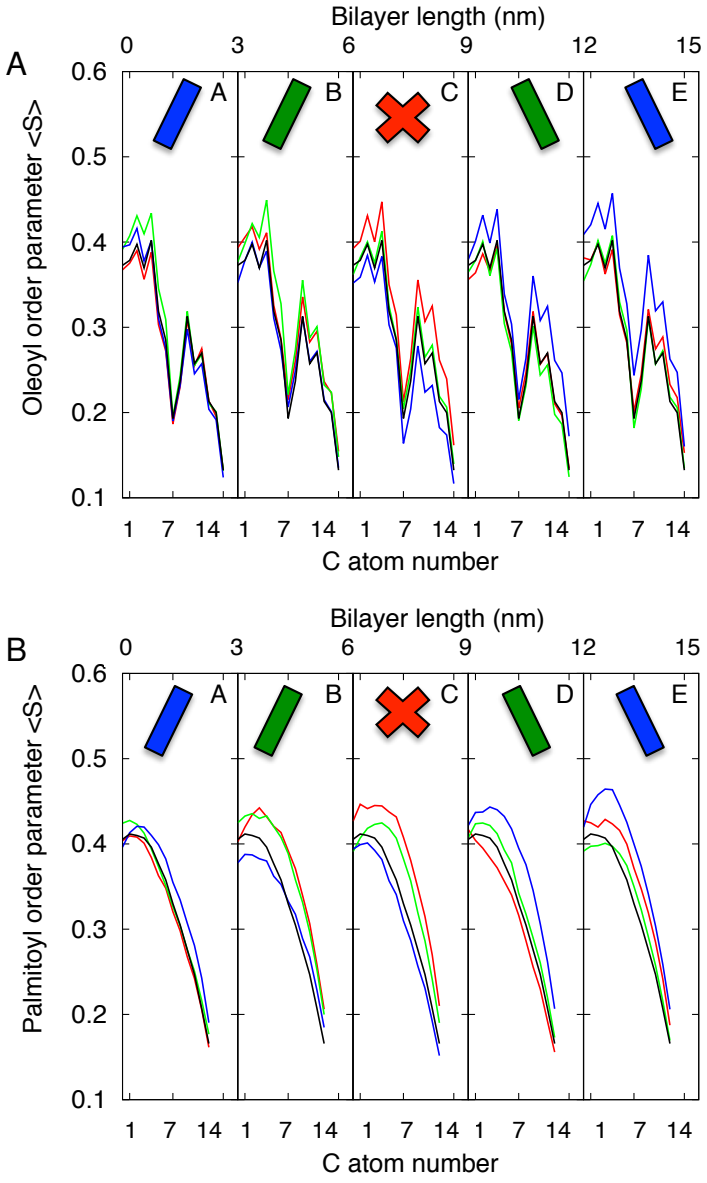

Figure 4.10: Average lipid order of (A) oleoyl tail, and (B) palmitoyl tail, from five regions of the bilayer (0 nm - 3 nm, 3 nm - 6 nm, 6 nm - 9 nm, and 12 nm - 15 nm as partitions A through to E, respectively) during simulations of Neu*$_{IV}$ as a monomer (blue), at an intermediate distance of approximately 4 nm between helices (green), and at the free energy global minimum (red). The position of each peptide is indicated by the corresponding marker. Lipid tail order calculations from a POPC bulk bilayer is shown in black. Standard deviation was calculated but the values were significantly small enough to exclude error bars.

is noted that the substitution produces a deep energy well upon peptide association in Neu*$_{IV}$.

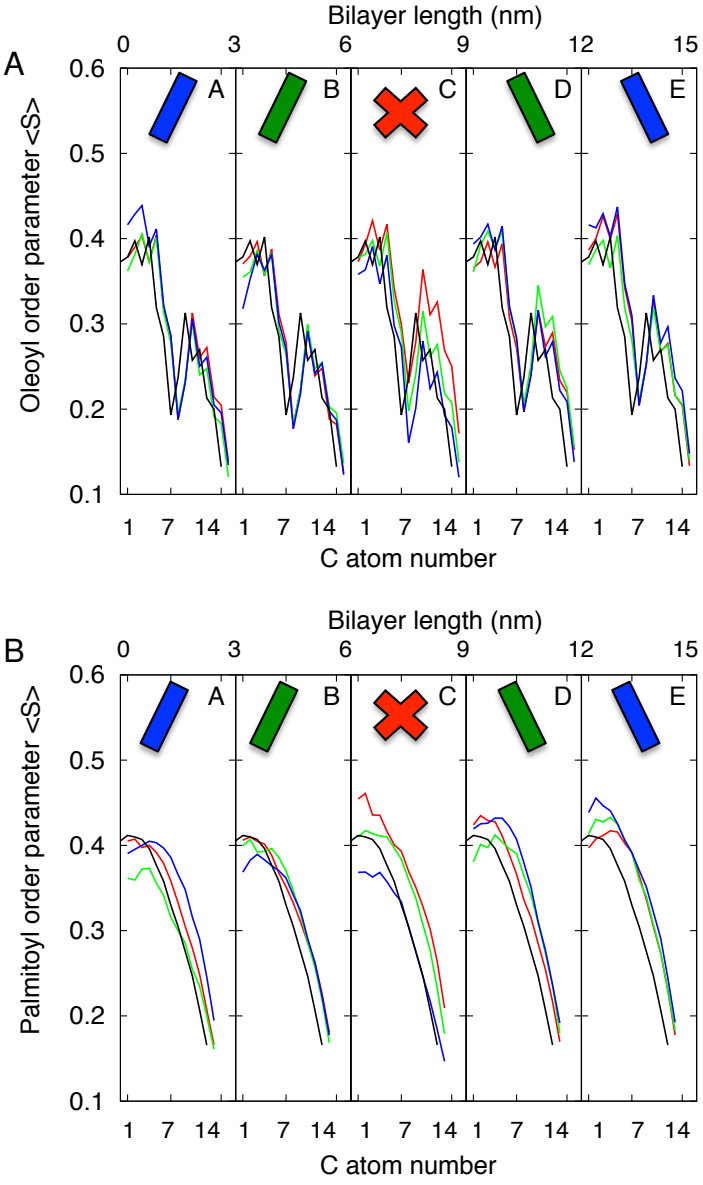

Figure 4.11: Average lipid order of (A) oleoyl tail, and (B) palmitoyl tail, from five regions of the bilayer (0 nm - 3 nm, 3 nm - 6 nm, 6 nm - 9 nm, and 12 nm - 15 nm as partitions A through to E, respectively) during simulations of Neu_{AG} as a monomer (blue), at an intermediate distance of approximately 4 nm between helices (green), and at the free energy global minimum (red). The position of each peptide is indicated by the corresponding marker. Lipid tail order calculations from a POPC bulk bilayer is shown in black. Standard deviation was calculated but the values were significantly small enough to exclude error bars.

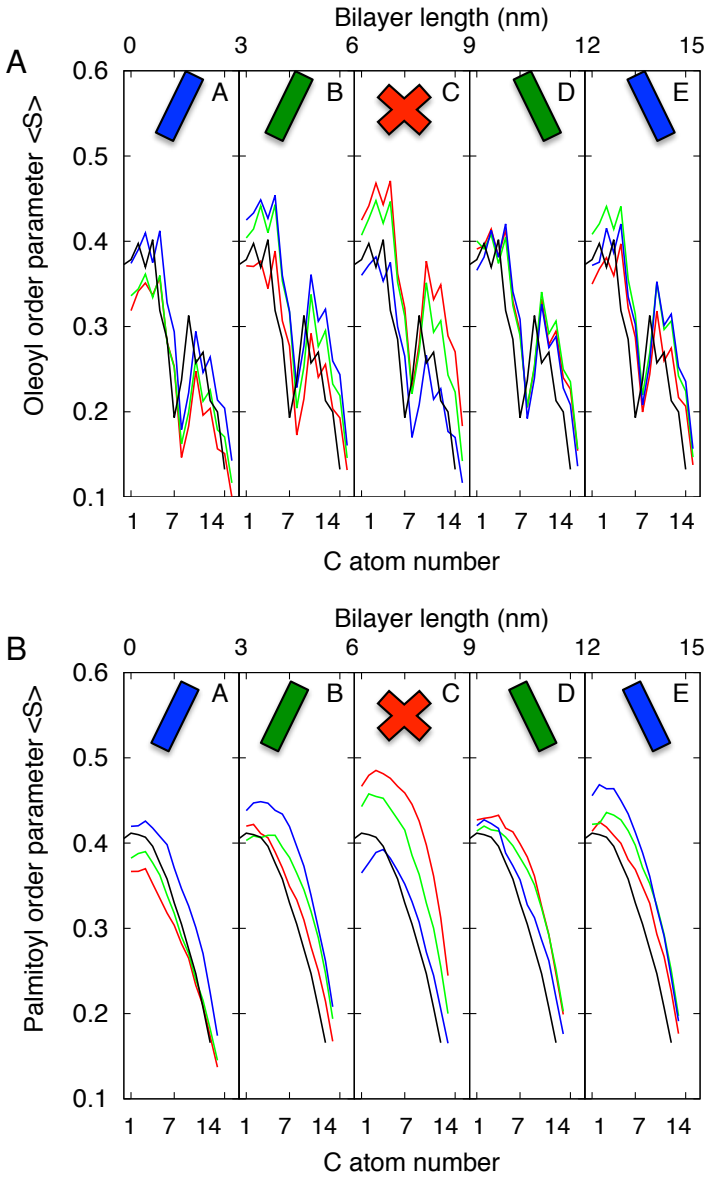

Figure 4.12: Average lipid order of (A) oleoyl tail, and (B) palmitoyl tail, from five regions of the bilayer (0 nm - 3 nm, 3 nm - 6 nm, 6 nm - 9 nm, and 12 nm - 15 nm as partitions A through to E, respectively) during simulations of Neu*$_{AG}$ as a monomer (blue), at an intermediate distance of approximately 4 nm between helices (green), and at the free energy global minimum (red). The position of each peptide is indicated by the corresponding marker. Lipid tail order calculations from a POPC bulk bilayer is shown in black. Standard deviation was calculated but the values were significantly small enough to exclude error bars.

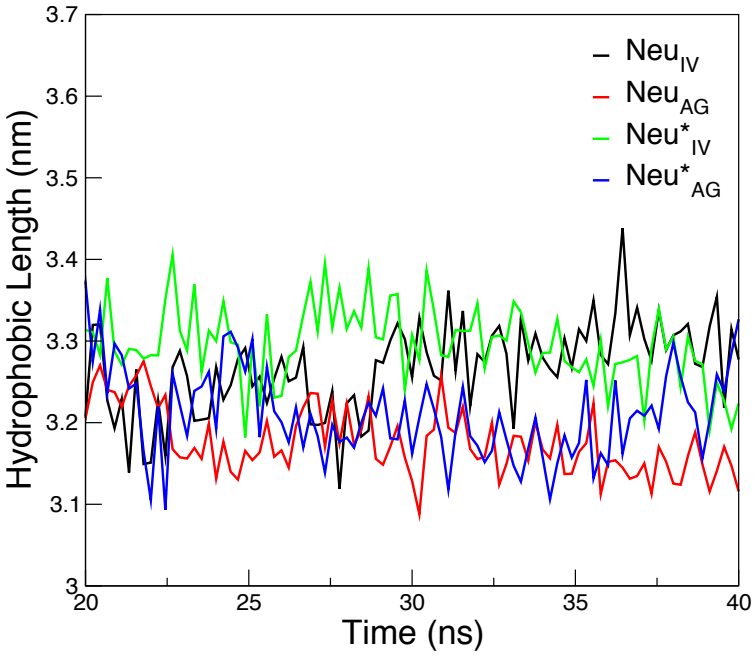

Figure 4.13: Average hydrophobic distance taken between F_{658} and I_{680} of both peptides of Neu_{IV}, Neu^*_{IV}, Neu_{AG} and Neu^*_{AG} taken over the last 20 ns of the trajectory at the global free energy minimum.

4.3.5 Dimer stabilisation through *inter*helical side chain interactions.

All four systems were found to be in favour of forming a dimer. By measuring the *xy*-lateral distance between helices, the umbrella window which approximated the location of the free energy minima was isolated. These trajectories were used for subsequent analysis over the last 20 ns.

The Neu_{IV} umbrella window whose *inter*helical distance was closest to 1.35 nm (Figure 4.16 (A)) was selected as the trajectory of the free energy minima. Bootstrap analysis suggested that the free energy values between 1.3 nm to 1.75 nm were statistical similar. Within this range, $\Delta G \approx -108$ kJ mol^{-1} was identified as an average free energy of self-association across this range. Residues directly in contact with the opposite helix and less than 10 Å (with the aim to identify residues immediately at the helix-helix interface) were identified using distance calculations in VMD (201) and are drawn in space fill representation in Figure 4.19 (A). Additional residues are still within short *inter*helical interaction distance, but do not form direct contacts at the forefront of the helix-helix interface. These additional residues can be identified from the *inter*helical contact maps in Figure 4.14 (A). At the interface there are no small-xxx-small motifs. The $I_{659}xxxV_{663}$ motif was within the short-range cut off but had rotated away from the interface. The foremost contacts are dominated by a collection of beta-branched side chain amino acids; phenylalanine, leucine and isoleucine, with side chain contacts spanning the length of both helices. Given that the umbrella potential had been applied between the centre-of-mass of each helix these beta-branched side chains help explain the location of the free energy minimum along the reaction coordinate. That is, the dominance of long side chains at the interfaces, increases the separation of the helices compared to what would be expected from a small-xxx-small motif.

The free energy minimum for Neu_{AG} was found within the range 0.9 to 1.19 nm. Bootstrap analysis suggests that the free energy values within this region are statistically similar. Within this range $\Delta G \approx -97$ kJ mol^{-1} was identified as an average free energy of self-association. The umbrella window whose trajectory approximated an average *inter*helical distance of 1.15 nm (Figure 4.16 (B)) was selected and is drawn Figure 4.19 (B). The value of the reaction coordinate at the free energy minimum is slightly less (helices are closer) than Neu_{IV} as may be the result of shorter side chain contacts. Although the $A_{661}xxxG_{665}$ motif was within the short-range cut off, it was not in direct contact with the opposite helix. In addition, side chains within the short-range cut off yet not

directly at the interface can be seen in the contact map Figure 4.14 (B). As with the Neu_{IV} dimer, there were no small-xxx-small motifs present and there were side chain contacts which spanned the entire length of each helix.

The free energy and *inter*helical side chain contacts for Neu^*_{IV} differ greatly from those of Neu_{IV}. The free energy minimum was $\Delta G \approx -55$kJ mol^{-1} on average and fell within a broad energy well statistically similar between the distances 1.3 nm and 1.8 nm. In comparison, this is 53 kJ mol^{-1} less favourable than Neu_{IV}. The free energy minima for Neu^*_{IV} and Neu_{IV} were found at a similar *inter*helical distance (see Figure 4.16 (C)). Unlike the wildtype, interface contacts were only found toward the N terminal of each helix with only nine residues in total directly involved in helix-helix contacts (Figure 4.20 (A). Interestingly, the *inter*helical packing of the $I_{659}xxxV_{663}$ motif was captured at the final frame of the free energy minimum trajectory, which contradicts the atom-atom distance plot in Figure 4.14 (B). Careful observation of the trajectory confirmed that the $I_{659}xxxV_{663}$ motif deviated from a $I_{659}xxxV_{663}$ to $I_{659}xxxV_{663}$ helix-helix potential interaction,and therefore it was not recorded as a more permanent *inter*helical feature on the atom-atom distance plot. The rearrangement of the helices is most likely down to the interation between the hydrophilic side chain of the glutamic acids and the hydrophobic lipid core. Both glutamic acids acid side chains orientated towards the hydrophobic lipid core. At least one of the glutamic acid side chains is within the distance needed to form a hydrogen bond between the carboxylic group on its side chain with the oxygen on the backbone of L_{660} (see Figure 4.17 (A)). A hydrogen bond would help elevate the energetic penalty incurred from positioning the glutamic acid side chain within the lipid core. Additional side chain contacts can be seen from the contact map in Figure 4.14 (B), yet there were no side chains within a potential *inter*helical contact distance at the C terminal half of both helices.

The Neu^*_{AG} umbrella window whose *inter*helical distance distribution approximated 0.9 nm (Figure 4.16 (D)) was selected as the best representational trajectory of the free energy minimum. The bootstrap analysis suggested that the free energy values between 0.77 nm to 1 nm were statistically similar. Therefore, within this region, $\Delta G \approx$ -58 kJ mol^{-1} was the average free energy of association. The wildtype was in favour compared to the mutant by 39 kJ mol^{-1}. The free energy of self-association between Neu^*_{IV} and Neu^*_{AG} was considered to be statistically similar although the minima for Neu^*_{AG} occurs at a smaller *inter*helical distance than for Neu^*_{IV}. Residues directly within contact of the opposite helix were identified and are drawn in Figure 4.20 (B). There are

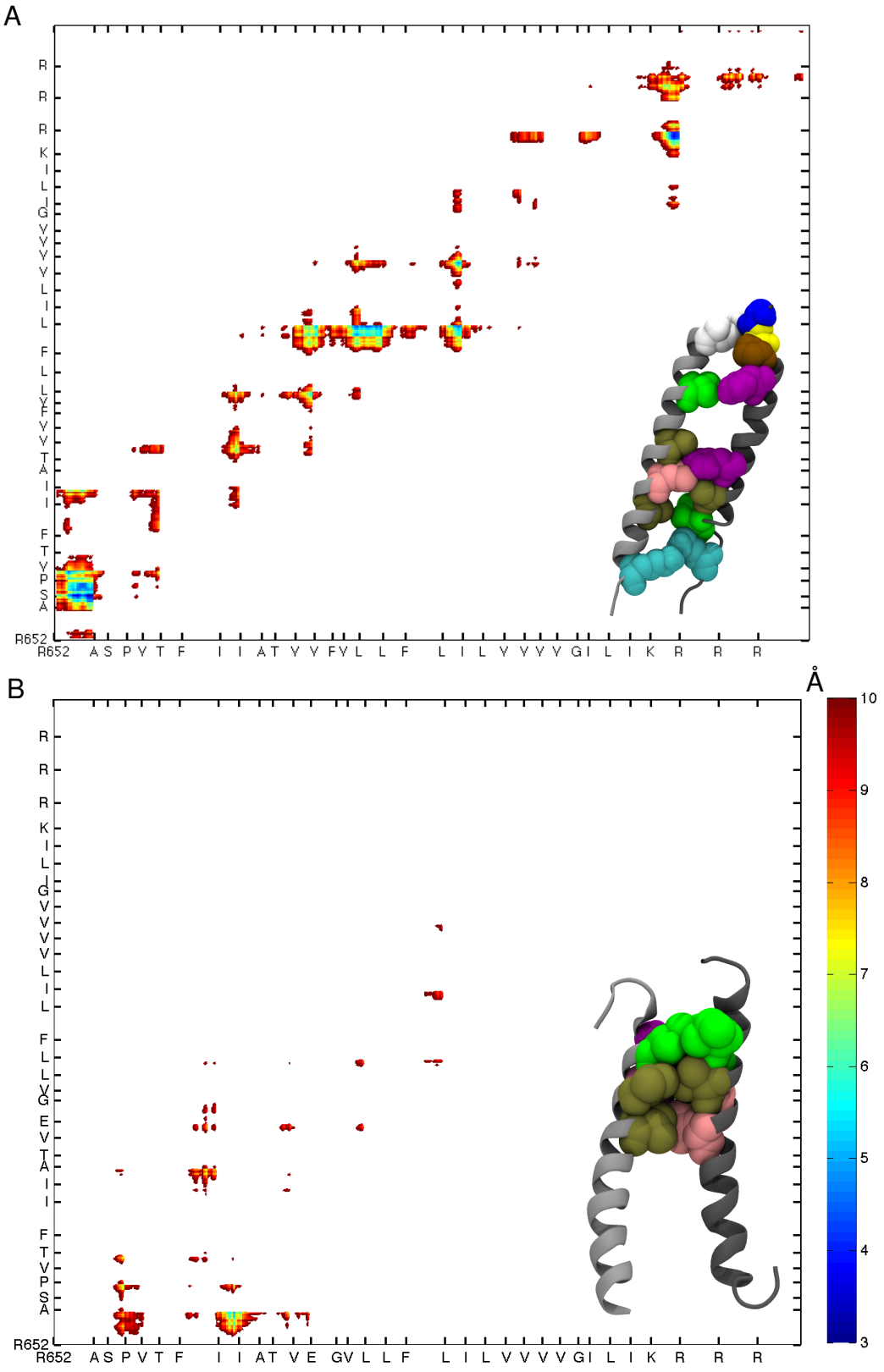

Figure 4.14: Interhelical atom-atom contact plots of (A) Neu_{IV}, and (B) Neu^*_{IV}. For clarity, distances beyond 10 Å were not included.

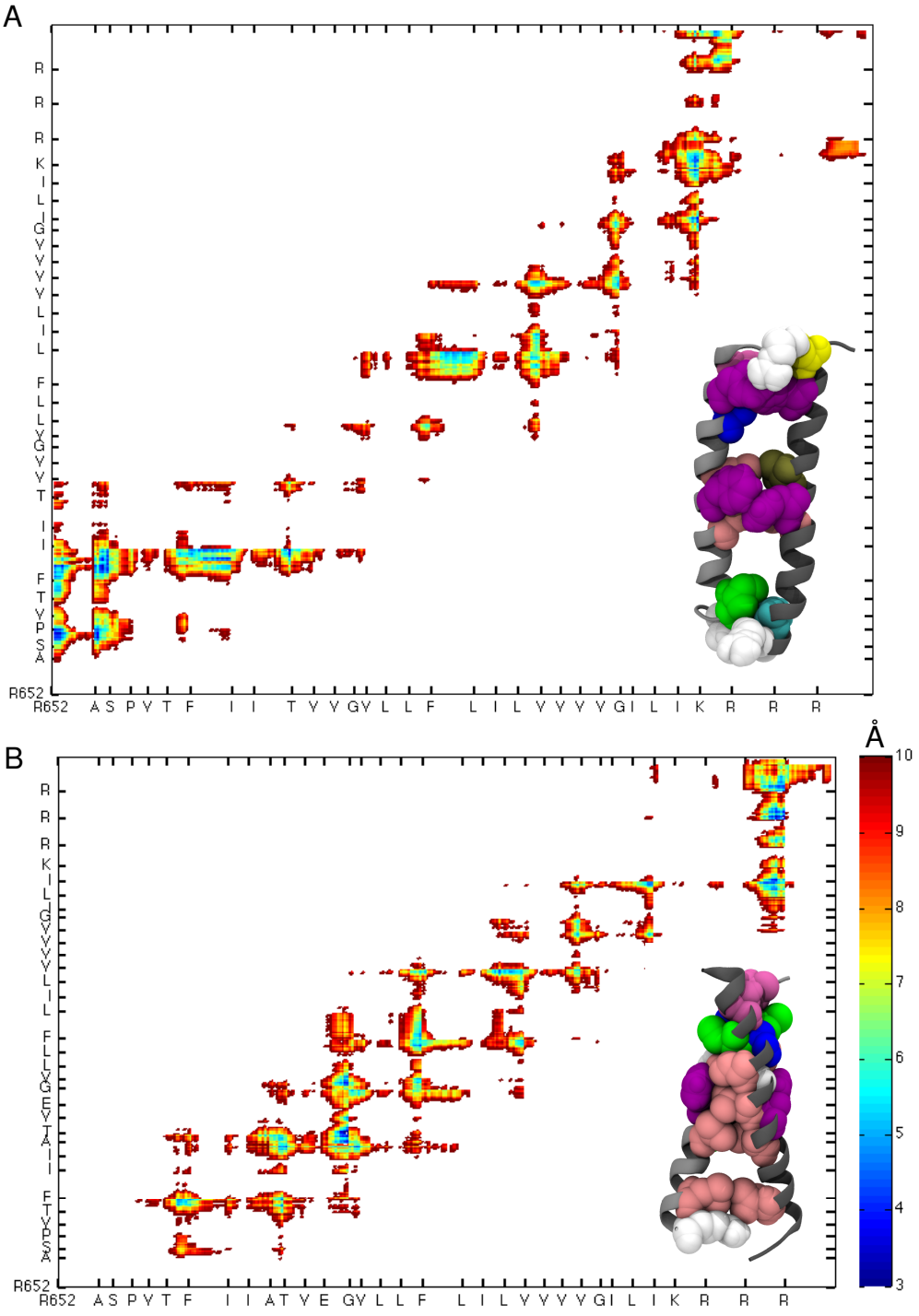

Figure 4.15: Interhelical atom-atom contact plots of (A) Neu_{AG}, and (B) Neu^*_{AG}. For clarity, distances beyond 10 Å were not included.

a significant number of contacts between side chains along the *inter*helical interface, spanning almost the entire length of both helices. The $A_{661}xxxG_{665}$ motif has been preserved and the E_{664} residue buries its side chain within the helix-helix interface (see Figure 4.17 (B)). All other potential contacts between residues were recorded in the contact map Figure 4.15 (B).

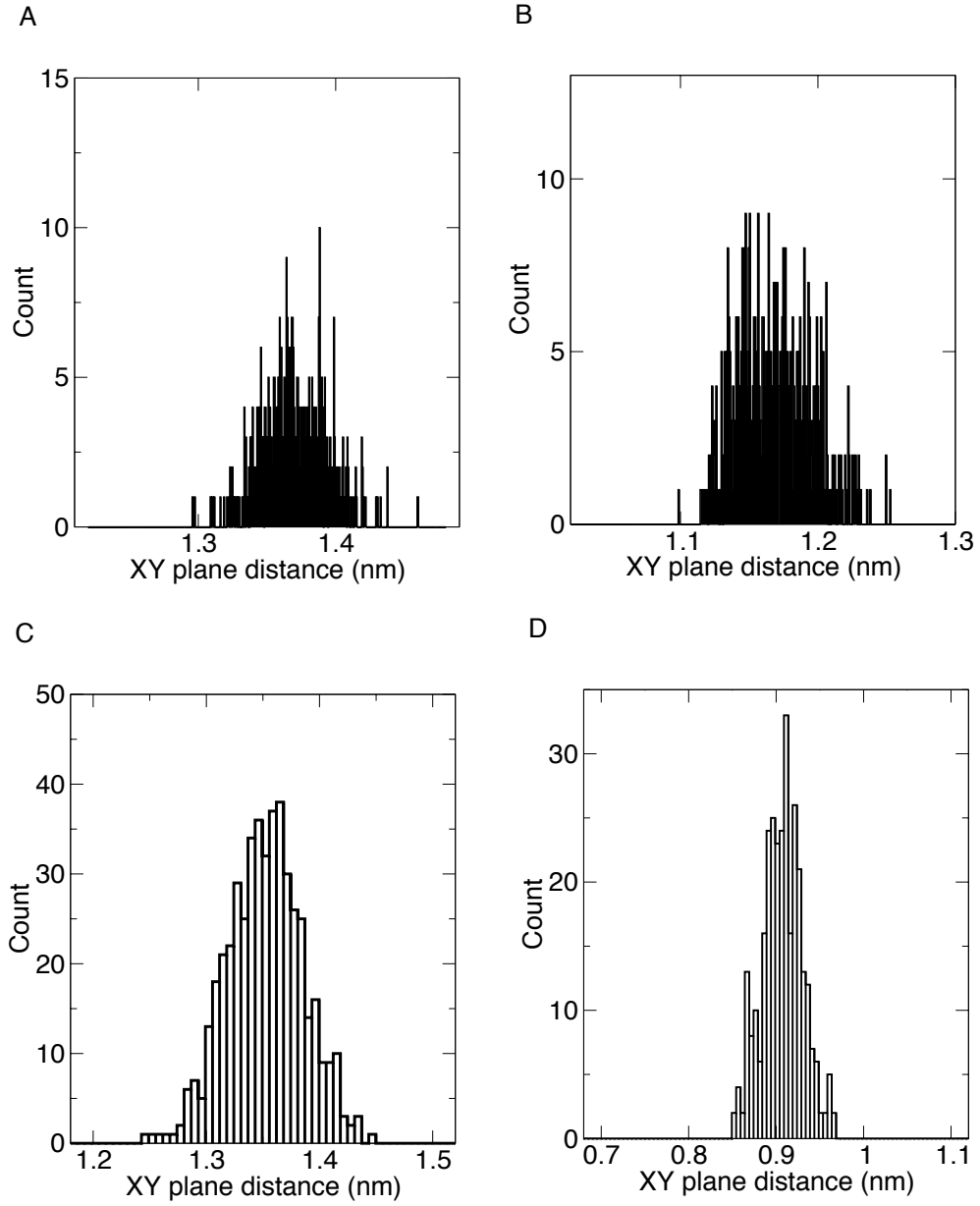

Figure 4.16: Interhelical distance distributions calculated along the *xy*-dimension. Taken from umbrella simulation windows pertaining to the free energy of systems of *(A)* Neu_{IV}, *(B)* Neu_{AG}, *(C)* Neu^*_{IV}, and *(D)* Neu^*_{AG} from simulation umbrella windows five, two, six, and zero, respectively.

All four global minimum trajectories were subjected to solvent accessible surface area analysis (SASA) to determine whether there was a significant change in solvent accessibility to each helix and to enable comparison with deuterium exchange experiments. Approximately, 14 nm^2 to 15 nm^2 of each dimer was exposed to the solvent (see Table 4.1), indicative of the polar residues at either terminus: arganine; serine; threonine; lysine and alanine, are partaking in the lipid-water interface. All four SASA calculations compare well to percent deuterium exchange detection of the Neu

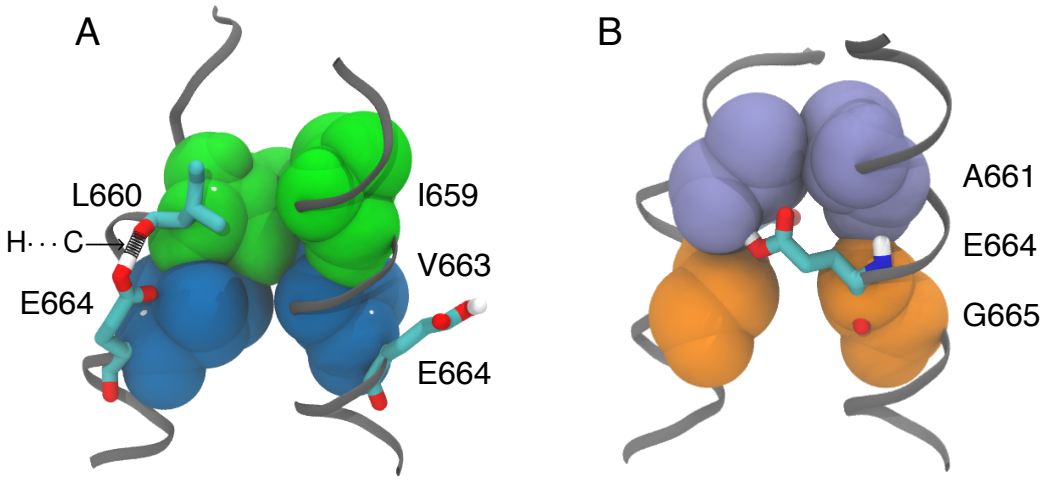

Figure 4.17: Final frame of the (A) Neu*$_{IV}$ $I_{659}xxxV_{663}$ and (B) Neu$_{AG}$ $A_{661}xxxG_{665}$, *inter*helical motifs in addition to E_{664}.

and Neu* peptide backbone which suggested that Neu and Neu* peptides were almost completely shielded from exchange by being buried inside the lipid membrane (46). There was particular interest in whether the glutamic acid substitution was interacting with water at the lipid-water interface as polar residues have been shown to mediate the permeation of water into the bilayer hydrophobic core (202, 203, 204), potentially disrupting helical stability. The density profiles of both Neu*$_{AG}$ and Neu*$_{IV}$ systems were calculated to determined whether there was a noticeable overlap between water and glutamic acid residues in either helix-orientation. As seen in Figure 4.18 (A), the glutamic acid residues (blue and violet) in the Neu*$_{AG}$ dimer do not overlap with the water. This was expected given that both residues are buried at the helix-helix interface. However, in the opposite helical orientation (Figure 4.18 (B)), one of the two glutamic acid residues in Neu*$_{IV}$ overlaps with the water density calculations.

It should be noted that the concentration of proto-oncogenic Neu wild type and oncogenic Neu dimers in POPC bilayers do not yield a selectable orientation of oligomerisation, therefore direct comparison to the $I_{659}xxxV_{663}$ and $A_{661}xxxG_{665}$ orientations presented in this study are not possible. Therefore, α-helical values were generated from the combined average of both $I_{659}xxxV_{663}$ and $A_{661}xxxG_{665}$ simulations using the Gromacs g_helix tool, as summarised in Table 4.1 , the α-helical composition of Neu$_{IV}$ was 57% ± 1% and Neu$_{AG}$ was 69% ± 1%. These results compare well to experimental CD results of proto-oncogenic Neu and oncogenic Neu in POPC bilayers, which yield an average α-helical composition of 67.5% and 63.2%, respectively (46).

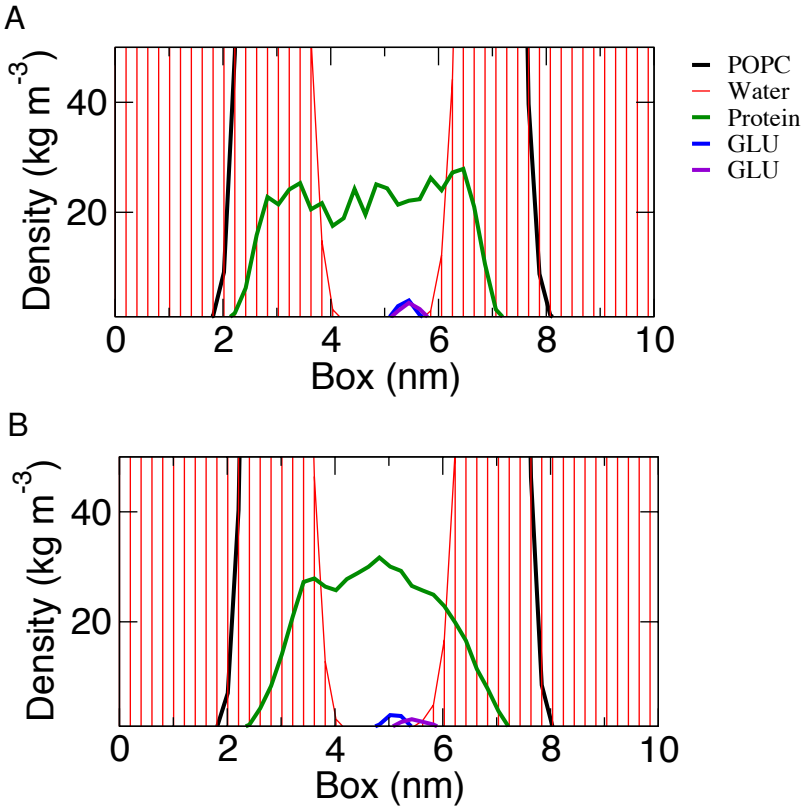

Figure 4.18: Particle density normal to the bilayer of (A) Neu*$_{AG}$, and (B) Neu*$_{IV}$. The presence of water is indicated by the red shaded segment.

The average dimer tilt angle at the global free energy minima is presented to enable comparisons with linear dichroism (LD) experimental measurements of tilt angle of proto-oncogenic Neu and oncogenic Neu in POPC bilayers (178). As summarised in Table 4.1, Neu$_{IV}$ and Neu$_{AG}$, with respect to the bilayer normal returned tilt angles of 45◦ ± 3° and 24° ± 2°, respectively, and the oncogenic form returned tilt angles of 54° ± 2° and 24° ± 4° for Neu*$_{IV}$ and Neu*$_{AG}$ dimer, respectively. LD experiments indicated that oncogenic Neu tilts in excess of 30°. The spectra of proto-oncogenic Neu suggested a tilt angle less than oncogenic Neu. As with the α-helical composition, a direct comparison to experimentation is not possible unless one considers an equal distribution of both orientations of oligomerisation across the population of Neu peptides. In such a case, the average dimer tilt angles in this study are 39° and 34.5° for oncogenic and proto-oncogenic Neu, respectively, which compares well to the experimental results.

4.4 Discussion

In this study, we investigated how a POPC model bilayer provides a mechanism for Neu TM domain dimerisation from the positive hydrophobic mismatch between the length of the peptides and the

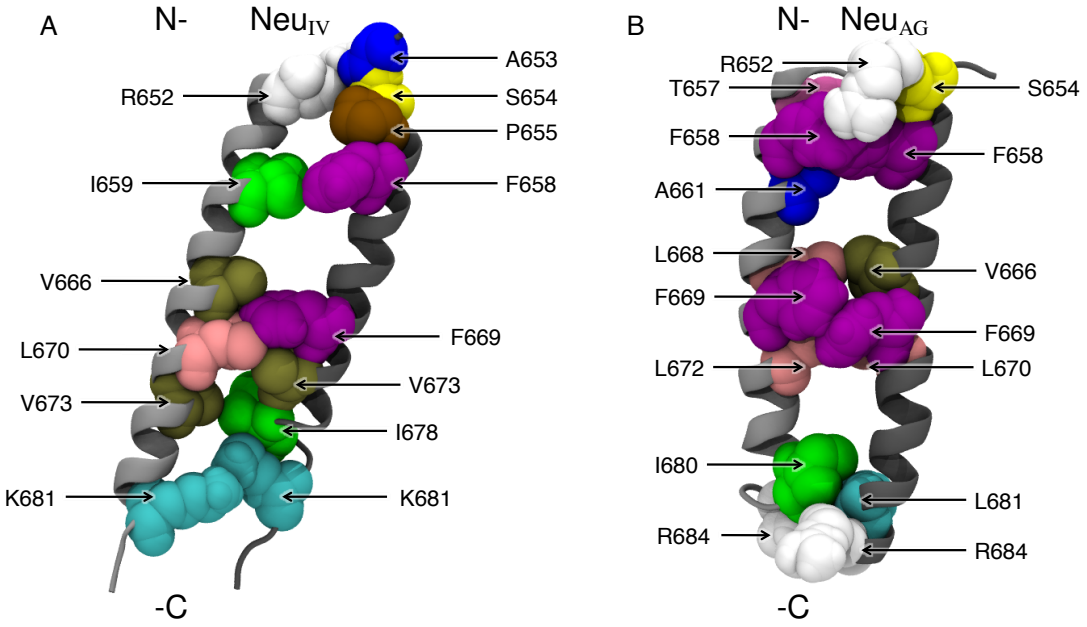

Figure 4.19: Final snapshots of dimers; (A) Neu_{IV}, and (B) Neu_{AG}, from the final structure found at the global free energy minimum. All residues that do not contribute to helix-helix packing at the interface were removed. The bilayer and water were removed for clarity. A helical ribbon represents the peptide backbone, and amino acids are represented by space filling spheres.

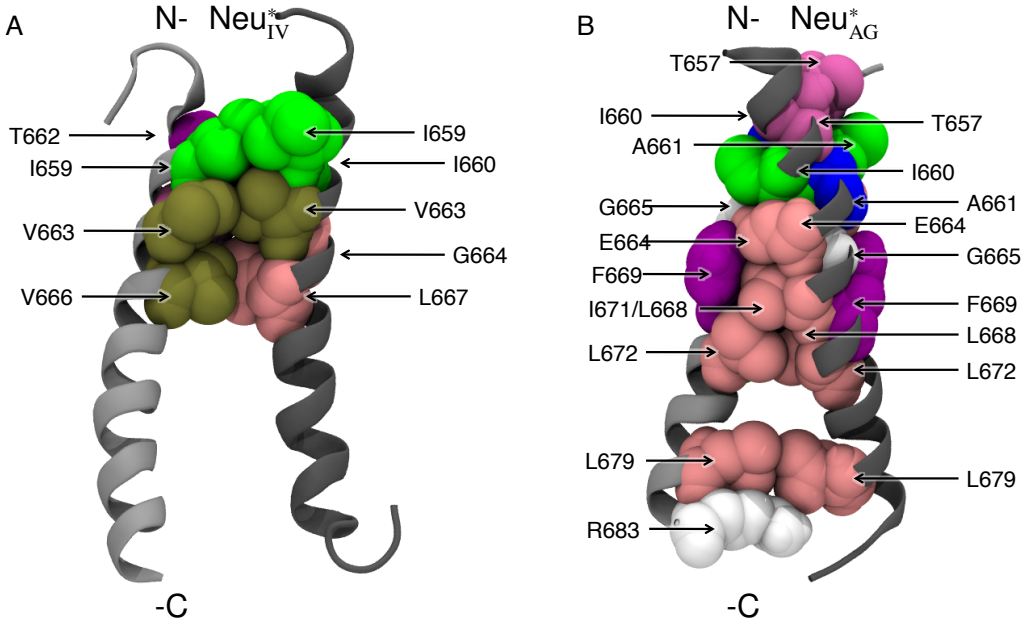

Figure 4.20: Final snapshots of dimers (A) NeuIV*, and (B) Neu^*_{AG} from the final structure found at the global free energy minimum. All residues that do not contribute to helix-helix packing at the interface were removed. The bilayer and water were removed for clarity. A helical ribbon represents the peptide backbone, and amino acids are represented by space filling spheres.

bilayer thickness. The concept of a top-down approach was initiated via the substitution of the single amino acid substitution, $V_{664}E$, focusing on the consequences to the lipid-mediated driving force and the differences to the helix-helix packing interface. To our knowledge, this has been the

	ΔG (kJ mol^{-1})	$\Delta\Delta G$ (kJ mol^{-1})	Tilt Angle (°)	Helicity (%)	SASA (nm^2)	Hydrophobic Length (nm)
Neu_{IV}	−108 ± 7	-	45 ± 5	57 ± 1	15 ± 1	3.22 ± 0.08
Neu_{AG}	−97 ± 10	-	24 ± 2	69 ± 1	14 ± 1	3.65 ± 0.07
Neu^*_{IV}	−55 ± 9	53 ± 2	54 ± 2	67 ± 2	14 ± 1	3.62 ± 0.07
Neu^*_{AG}	−58 ± 8	39 ± 2	24 ± 4	60 ± 1	14 ± 1	3.18 ± 0.06

Table 4.1: Free energy and structural calculations of Neu and oncogenic Neu dimers. The $\Delta\Delta G$ is between the oncogenic and proto-oncogenic sharing the same interface. Standard deviation was used for estimating all errors.

first study involving the calculation of the self-association Neu and its oncogenic form over such a long reaction coordinate whilst taking into account the helix-helix orientation. To further this study, it would be insightful to employ the use of meta-dynamics as a function of dimer rotation to better understand the mode of receptor activation.

4.4.1 The Neu and Neu* dimers are stabilised by side chain interactions.

The noticeable difference in free energy minima between the two helical orientations of Neu and Neu* is likely to arise from the differences in side chain interactions between homo dimers. One of the first proposed models of oncogenic kinase activation and increased dimerisation (72, 74) was based on the protonation of the ionized glutamic side chain due to the pH of the membrane interior. As a result, the E_{664} side chain would stabilise receptor dimerisation by *inter*helical hydrogen bonding of its carboxyl group with the carbonyl oxygen of A_{661}. The size of G_{665}, on the same face as A_{661} and E_{664} would allow for close packing. Yet, NMR studies were unable to confirm direct evidence of hydrogen bonding as proposed in the model (73) although model peptides had been solubilised in TFE solution, providing a greater degree of freedom unlike the confines of a membrane bilayer. On the other hand, our trajectories obtained from the oncogenic Neu^*_{AG} free energy minimum presented a tightly coupled helix-helix interface, with $A_{664}xxxG_{665}$ packing at the interface. At this close distance, the protonated side chain of E_{664} was within a very close distance (< 4 Å) of the backbone of A_{661}, suggesting the potential for *inter*helical hydrogen bonding. In the proto-oncogenic Neu_{AG}, the free energy minimum is an additional 5 Å to 10 Å along the reaction coordinate compared with the structure of Neu^*_{AG}. This supports the idea that a close interaction is required to shield the polarity of the protonated glutamic side chains. Our results are supported further from pK_a measurements and ^{13}C chemical shift anisotropy of E_{664} in oncogenic Neu (205), yielding results that suggest that protonation of E_{664} allows a helical conformation followed by a partitioning into the membrane and finally dimerisation. The helical orientations of

Neu*$_{IV}$ dimer results in the glutamic acid on each helix being exposed to the hydrophobic lipid environment. Our data suggests that the energetic penalty for this hydrophobic mismatch is elevated through a hydrogen bond formation of the glutamic acid closest to the lipid core (Figure 4.17 (A)), whilst maintaining a smaller dimer tilt angle than the Neu*$_{A}$ dimer to encourage the interaction between water and lipid head groups with the second glutamic acid side chain (Figure 4.18 (B)). This mechanism is not seen in Neu*AG given the packing of the glutamic acid side chains at the helix-helix interface.

A conformation change to the helical backbone, a subsequence activation to the intracellular tyrosine kinase domain, has been suggested as an alternative model by Brandt-Rauf *et al.,* (206, 207) who hypothesised that the backbone immediate to the E_{664} site would dramatically kink and disrupt the α-helical conformation. However, three-dimensional structures of proto-oncogenic Neu and oncogenic Neu obtained by NMR spectroscopy (73) confirmed that the overall α-helical structure was preserved. Further more, in neither of the Neu*$_{AG}$ or Neu*$_{IV}$ global free energy minimum trajectories was a dramatic change in α-helical composition or kink to the backbone observed. The levels of α-helicity over both $I_{659}xxxV_{663}$ and $A_{661}xxxG_{665}$ respective forms of proto- and oncogenic Neu from the free energy minimum trajectories match well with CD experiments (46). NMR data of the two forms of Neu in detergent micelles also supports our reported helical values (208): even though there was a significant difference in chemical shift between proto-oncogenic and oncogenic Neu in DPC, the data was consistent with a helical structure in residues V_{656} through I_{680}.

Although both Neu$_{IV}$ and Neu$_{AG}$ form a dimer without either $I_{659}xxxV_{665}$ or $A_{661}xxxG_{665}$ motifs directly at the helix-helix interface, it should not be ruled out that they are essential for dimerisation as seen in the works of Beevers *et al.,* (46). The PMF profiles captured in this study suggest a lipid mediated driving force powered by the requirements to minimise the ordered lipids in the simulation cell. By forming a dimer, the number of annular lipids is greatly reduced (given that one face of each helix is involved in helix-helix packing rather than helix-lipid interaction). In a system comprised of a whole host of lipid types, such pressure between lipid tail length and peptide hydrophobic span may not play such a significant role, and thus a stable dimer may rely more on the energetic contributions of its interacting side chains and the interactions between the side chains and native biomolecules, such as cholesterol (209).

4.4.2 Interhelical approach is a lipid-mediated process.

At 33 amino acids, the sequence used to represent the TM domain of Neu is very long. Compensation for the length with a large tilt angle is evidently not enough, and the hydrophobic mismatch between peptide and bilayer result in an adjustment to the bilayer thickness. This is seen from each of the cross sectional bilayer thickness plots with respect to the location of the dimer.

A possible consequence of the ordering of lipids around the helices is that dimerisation is driven by the structural rearrangement of lipids surrounding the peptides. As the two lipid-mediated *zones of influence* intersect, the free energy begins a descent. During the descent, the bilayer thickness goes from two distinct thicker regions, through a region of intermediate thickness, to finally a broad single thicker region to accommodate the dimer. As the peptides approach, a greater proportion of lipids return to the bulk liquid crystalline phase. As a consequence, the increase in lipid order follows the location of the peptides, only to resume a disordered state in the wake of the peptide. This process suggests an increase in conformational entropy of the lipids as the peptides associate (210). Finally, as both peptides form a dimer, the surface area available for peptide-lipid contact, and subsequent ordering of the alkyl lipid tails, is reduced due to direct peptide-peptide contact, thereby decreasing the overall order in the system.

Across all four systems, the free energy plateau requires over 5 nm of *inter*helical separation (around 7 nm along the reaction coordinate) before the difference in free energy becomes independent of peptide-peptide distance. Each peptide can then be considered a true monomer, beyond all non-bonding interactions, and any lipid mediated interactions. After observation of the umbrella window trajectories furthest along the reaction coordinate, and a lateral calculation of average bilayer thickness, a region of the bilayer between peptides returned to a bulk lipid thickness.

The change between the oncogenic and proto-oncogenic PMF profiles is most likely a result of introducing a hydrophilic residue into the bilayer core. As such we would expect to see contention between the position of the residue and the hydrophobic mismatch with the lipid tails. The $V_{664}E$ substitution in the human ErbB2 ortholog of rat Neu has been shown to change the membrane boundary of the TM helix by inducing a transverse shift (a shift in the dimer position, along the dimension normal to the bilayer). This has been confirmed using an alternating leu-ala sequence (174), suggesting that the transverse shift is not strongly affected by the neighbouring

hydrophobic residues (211). Although the distance between E_{664} and the average hydrophilic head group was not measured, a transverse shift would help explain the differences between oncogenic and proto-oncogenic Neu PMF profiles. This is supported further by the orientation of E_{664} during Neu^*_{IV} dimerisation. As observed in Figure 4.17 (A), due to the excessive tilt angle ($54° \pm 2°$), the E_{664} closest to the lipid-water interface could induce a slight transverse shift so as to satisfy the hydrophobic mismatch. As a result, the E_{664} now situated on the lipid-core side of the dimer has a reduced chance of interacting with the lipid head group region and potentially forms a hydrogen bond with the backbone oxygen atom of L_{660} on the same helix (Figure 4.17 (B)).

The monomeric tilt angles between helical orientations ($I_{659}xxxV_{663}$ and $A_{661}xxxG_{665}$) of at least one peptide from either proto-oncogenic or oncogenic Neu were similar, suggesting that the same sequence regardless of helical orientation had adopted a monomeric state. It is worth noting that the tilt angle of oncogenic and proto-oncogenic peptides, 4 nm along the reaction coordinate, were significantly different according to their helix-helix orientation. Neu_{AG} was shown to adopt a greater tilt angle than Neu_{IV}, which clearly demonstrated an elevation to a significantly greater portion of the bilayer than that of the Neu_{IV} simulation. This helps explain the greater free energy value of Neu_{AG} with respect to Neu_{IV}. Unfortunately, this mechanism cannot explain the switch in free energy order between the oncogenic Neu orientations.

4.4.3 Both Neu helical orientations have a similar propensity for dimerisation.

The results show that Neu_{IV} is energetically more favourable than Neu_{AG}, although statistical uncertainty suggests that there is equal drive towards dimerisation across some of the free energy descent. Although the Neu_{IV} orientation maintains a lower free energy average, upon dimerisation, the statistical uncertainty suggests equal propensity for stabilising the dimer. Some studies indicate that the membrane mimetic environment can influence the propensity for Neu dimerisation. For example, Neu is thought to be monomeric in detergent micelles but forms a dimer in bicelles (212). Other studies have given semi-quantitative approximations of dimer propensity between different forms of Neu using an *E. coli* inner membrane (46). In such a case, one must consider a large number of unknown variables going into these calculations, including the variation in: lipid composition; targeted oligomeric state; and interaction with an uncountable number of host proteins. Activation of the cytoplasmic kinase domains requires a conformational change in the TM domain region. As

a switch with as yet no identifiable ligand, it is thought to activate upon autophosphorylation from a conformal change leading into the kinase domain (213). A change in propensity between the two orientations of oncogenic Neu as described in the PMF profiles presented here, may point towards the unregulated cell growth that is well established in the literature (214).

4.4.4 Both Neu* helical orientations have a similar free energy values, yet dimerisation of Neu*$_{AG}$ and Neu*$_{IV}$ may halt along a free energy plateau.

Substituting V_{665} for a protonated glutamic acid made a noticeable change to the free energy profile. The difference in free energy between the proto-oncogenic Neu and the oncogenic form had been reduced; although still in favour of dimerisation, the propensity was less. In addition, Neu*$_{AG}$ becomes more energetically favourable than Neu*$_{IV}$, which is not the case for Neu. Yet, unlike Neu*$_{IV}$, upon self-association Neu*$_{AG}$ looses its drive toward dimerisation just before the short-range interactions due to a significantly long free energy plateau. This is also observed for Neu*$_{IV}$ from 4 nm through 6 nm along the reaction coordinate. The PMF profiles imply that a switch in helix-helix orientation as the monomers approach each other would drive Neu* into a dimer with $I_{659}xxxV_{663}$ at the interface. Conversely, the results indicate that if dimerisation occurred with a $A_{661}xxxG_{665}$ helix-helix orientation, the deep global free energy well may prevent Neu* from exploring a different conformation, as would be the case from the statistically equivalent free energy orientations of Neu. The mechanism proposed here of a locked helix-helix orientation of Neu*, support the increase in activation propensity of Neu* in CHO (Chinese hamster ovary) cells (215). It is worth noting that a recent publication by Prasanna *et al.,* (216) explored the PMF profiles of ErbB2, including the mutant $V_{664}E$ in addition to the perturbation to the annual lipids using the MARTINI CG force field. They report similar changes to the bilayer thickness, however, as observed in chapter 6, the early convergence of the PMF profile maybe as a result of the force field. Further more, dimers were formed using self-assembly with little consideration to the helix-helix orientation.

4.5 Conclusion

Given that these calculations are computationally expensive, it has become commonplace to define a reaction coordinate no further than the cut off of the short-range interaction potential. We have

shown the profound affect that long-range interactions, mediated through long-ranged lipid-lipid interactions, have on the PMF profile of self-association of the receptor tyrosine kinase Neu along a *together-apart-together* reaction coordinate. By doing so, it was revealed that the positive hydrophobic mismatch between the hydrophobic length of the Neu TM domain and the hydrophobic length of the POPC tails was acting as a driving force by minimising the number of ordered tails in the system. Interestingly, although this mechanism was still observed in the oncogenic form, the PMF profile was not as pronounced as in the proto-oncogenic form. We observed a potential switch in preferred helix-helix oligomerisation between the oncogenic and proto-oncogenic form. As a dimer, the interaction motifs $I_{659}xxxV_{663}$ and $A_{661}xxxG_{665}$ were not directly involved with helix-helix packing in the oncogenic dimer. However, they were present in both proto-oncogenic helix-helix orientations. A mechanism of shielding the polarity of E_{664} was observed in the form of burying both residues at the helix-helix interface, or forming *intra*helical hydrogen bonds with one E_{664} whilst the second E_{664} is positioned towards the water-lipid interface. Finally, we suggest that the difference between PMF profiles of both helical orientations from the same sequence is sequence independent, and rather a result of a complex mechanism of tilt angle, bilayer thickness, lipid order, and glutamic acid location (in the case of oncogenic Neu). Further investigation would be required to unravel this mechanism.

This study has not explicitly addressed a rotational mechanism of tyrosine kinase activation (179), however, it does suggest that lateral association between Neu peptides is influenced greatly by the hydrophobic environment and the $V_{664}E$ substitution.

Chapter 5

Self association of low complexity scaffold TM domains in a biological environment using the *in vivo* GALLEX assay

5.1 Introduction

Over two decades of research into TM domain oligomerisation suggests that one of the key components thought to stabilise TM domain lateral interactions in the two stage model are interactions motifs (217). Working from a native TM sequence enforces a *top-down* approach, usually through alanine mutations of key amino acids. Unfortunately, the remaining unaltered TM sequence may continue to exert unaccounted for influences on the tertiary protein fold. This chapter looks at combining the well-known interaction motifs, GG4, heptad repeat, and single polar residues, on a low sequence complexity scaffold. Starting at a low sequence complexity heptad repeat, energetic contributions captured by introducing simple interaction motifs in a biological membrane (33). This *bottom-up* approach introduces a rational design through the arrangement of interaction motifs as *building blocks*.

The experimental work reported in this chapter aims to characterise the order of relative strength

of homo-dimerisation between TM domains of low sequence complexity scaffolds in a biological membrane, and by extension report the apparent free energy of self-association. Two sets of low sequence complexity, heptad repeat scaffolds were complemented by different helix-helix interaction motifs. Genes encoding these sequences were ligated into the pBLM100 destination vector for the GALLEX assay and expressed with IPTG. Chimeric proteins were tested for correct orientation across the inner membrane using the MalE complementation assay, and in some cases a sodium hydroxide wash was performed to determine if the protein chimera was associating with the inner membrane. The TM domain sequence along with three flanking residues from both linker regions of the chimera protein were theoretically assessed for their propensity of insertion via the endoplasmic reticulum Sec 61 translocon (218) using the online tool ΔGPredictor.

The order of relative strength of homo-dimerisation between low complexity scaffolds was determined using the GALLEX assay. The ΔG^{app} of self-association of scaffolds L17 and AZ2, along with their respective mutant sequences: L17(GG4); L17L9Q; AZ2$(GG4)_L$; and AZ2$(GG4)_A$ were calculated by varying the concentration of IPTG as described in section 3.8.2. The propensity for self-association and relevant free energy calculations of each low complexity scaffold are summarised in the discussion.

5.1.1 Low complexity scaffold design

The rational design for interaction motif building blocks, as illustrated in Figure 5.1, begins with a low sequence complexity heptad repeat. The first of two levels of complexity is introduced by adding a single interaction motif: be it a GG4 or a single polar residue. The second level, a single point mutation to the interaction motifs, aims to disrupt the interaction motif.

Each low complexity scaffold consisted of seventeen amino acids thought to adopt a left-handed, positive crossing angle (also known as a left-handed coiled-coil). The first design, polyleucine (L17), an α-helix rich in leucine, has typically been shown to conform to a heptad repeat pattern and adopt a left-handed crossing angle as a dimer (54). Residues L_3, L_6, L_{10}, L_{13}, and L_{17} make up the primary helix-helix interface (see L17, Table 5.1). The second low complexity scaffold, the alanine zipper (AZ2), consists of residues A_1, A_{4-5}, A_8, A_{11-12}, and A_{15}, and residues L_{2-3}, L_{6-7}, L_{9-10}, L_{13-14}, and L_{16-17} (see AZ2, Table 5.1). By itself alanine was found to either marginally exceed the hydrophobic threshold required for a stable membrane α-helix (219), or have a hydrophobicity

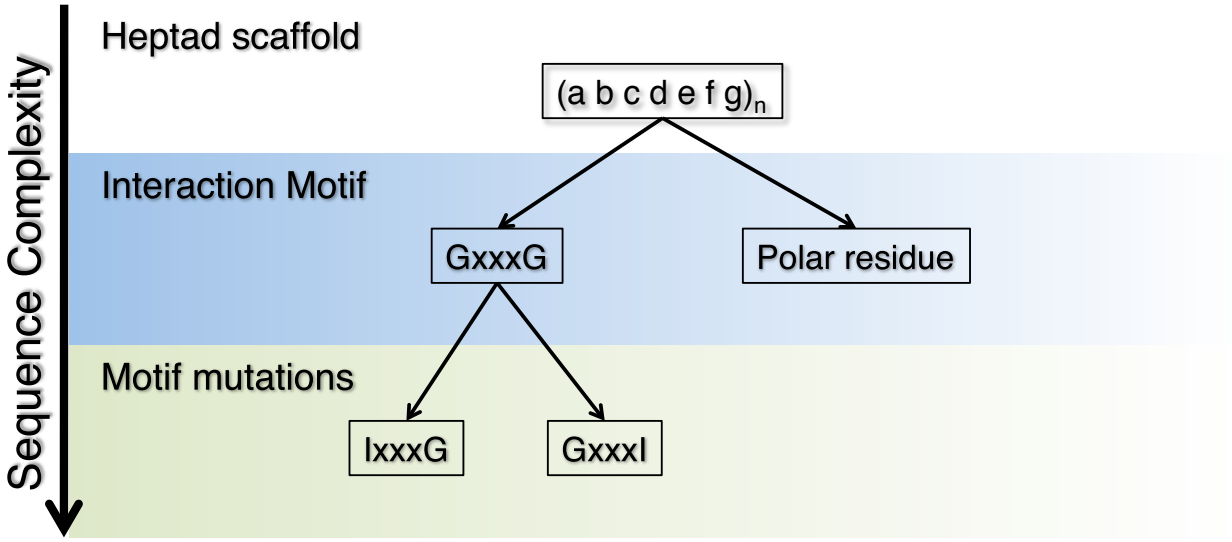

Figure 5.1: Rationale behind the *bottom-up* approach to low sequence complexity building blocks. From a heptad repeat sequence, either a GG4 or polar residue is introduced. Finally, the GG4 is disrupted from a single point mutation.

scale bordering a stable solvent α-helix (220), yet a mixed leucine-alanine peptide was identified as forming a stable integral membrane α-helix (221, 222). According to the early TOXR work by Gurezka *et al.* (106), the primary homo-dimer interface of an alanine zipper contained leucine at heptad repeat positions *a* and *d*. They found that a single point mutation of either site to alanine resulted in significant disruption to the dimer.

5.1.2 Polyleucine scaffold design

This study presents three individual modifications to the low complexity L17 scaffold. The first modification was the inclusion of a GG4 motif central to the hydrophobic TM domain (see L17(GG4), Table 5.1). In the second modification, the second glycine in the GG4 interaction motif was substituted for an isoleucine (see L17(GG4)G11I, Table 5.1). The GG4 interaction motif has shown to contribute significantly to helix-helix oligomerisation of TM domains (97, 24). Given the disruptive effect of the G83I substitution in the Glycophorin A dimer, any GG4 driven association of the polyleucine sequence should be disrupted. The third modification was the incorporation of a single glutamine residue at the centre of the TM domain (see L17L9Q, Table 5.1). Polar residues have shown to contribute significantly to helix-helix oligomerisation of TM domains (79, 69, 70) and are an evolutionary conserved feature of many native TM proteins (223).

5.1.3 Alanine zipper scaffold design

The AZ2 coiled coil heptad repeat sequence is thought to adopt a small left-handed crossing angle (224). The substitutions L_6G and $L_{10}G$ were made to introduce a GG4 interaction motif with the aim of further stabilising the oligomerisation of AZ2 (see $AZ2(GG4)_L$, Table 5.1). Alternatively, it may introduce contention between the two motifs as GG4 is found predominately in right-handed dimers (225). GG4 occurs frequently and in association with β-branched valine and isoleucine (28), the length of leucine's γ-branched side chain may prevent a GG4 helix-helix orientation from maximising its packing surface. The work of Lear *et al*., (225) found an improvement to self-association of a membrane-soluble peptide coiled coils with glycine at positions *a* and *d* over two heptads. It must be noted that their model peptide was more complex in sequence than AZ2 and featured an extra GG4 motif.

The next set of modifications, denoted $AZ2(GG4)_LG6I$ and $AZ2(GG4)_LG11I$ (see Table 5.1) introduces a single isoleucine in place of a glycine. These two scaffolds will help decipher whether GG4 is driving self-association, or if an additional interface could stabilise self-association during disruption to the GG4 on the leucine face.

Finally, a serine residue is incorporated at two different locations of the AZ2 sequence to determine whether serine can encourage self-association of an AZ2 low complexity scaffold. Serine residues replaced the leucine primary binding residues L_6 and L_{10} in sequences AZ2L6S and AZ2L10S (see Table 5.1), respectively. Calculations by Sal-Man *et al.,* (226) estimated that motifs of multiple serine residues could contribute up to 30% to the free energy of association.

5.1.4 Bifunctional alanine zipper

The low complexity scaffold $AZ2(GG4)_A$ was the result of the A_8G and $A_{12}G$ substitutions (see Table 5.1). Unlike $AZ2(GG4)_L$ where contention was purposely introduced to what was assumed to be the primary binding site, $AZ2(GG4)_A$ was intended to introduce a secondary helix-helix binding site. The additional modifications to $AZ2(GG4)_A$: $AZ2(GG4)_AG8I$, $AZ2(GG4)_AG12I$, $AZ2(GG4)_AL6A$ and $AZ2(GG4)_AL10A$ would determine whether a dimer could switch between binding sites.

Heptad template	f_1	g_2	a_3	b_4	c_5	d_6	e_7	f_8	g_9	a_{10}	b_{11}	c_{12}	d_{13}	e_{14}	f_{15}	g_{16}	a_{17}
L17	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
AZ2	A	L	L	A	A	L	L	A	L	L	A	A	L	L	A	L	L
L17	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
L17(GG4)	L	L	L	L	L	L	G	L	L	L	G	L	L	L	L	L	L
L17(GG4)G11I	L	L	L	L	L	L	G	L	L	L	I	L	L	L	L	L	L
L17L9Q	L	L	L	L	L	L	L	L	Q	L	L	L	L	L	L	L	L
AZ2	A	L	L	A	A	L	L	A	L	L	A	A	L	L	A	L	L
AZ2L6S	A	L	L	A	A	S	L	A	L	L	A	A	L	L	A	L	L
AZ2L10S	A	L	L	A	A	L	L	A	L	S	A	A	L	L	A	L	L
AZ2$(GG4)_L$	A	L	L	A	A	G	L	A	L	G	A	A	L	L	A	L	L
AZ2$(GG4)_L$G6I	A	L	L	A	A	I	L	A	L	G	A	A	L	L	A	L	L
AZ2$(GG4)_L$G10I	A	L	L	A	A	G	L	A	L	I	A	A	L	L	A	L	L
AZ2$(GG4)_A$	A	L	L	A	A	L	L	G	L	L	A	G	L	L	A	L	L
AZ2$(GG4)_A$G8I	A	L	L	A	A	L	L	I	L	L	A	G	L	L	A	L	L
AZ2$(GG4)_A$G12I	A	L	L	A	A	L	L	G	L	L	A	I	L	L	A	L	L
AZ2$(GG4)_A$L6A	A	L	L	A	A	A	L	G	L	L	A	G	L	L	A	L	L
AZ2$(GG4)_A$L10A	A	L	L	A	A	L	L	G	L	A	A	G	L	L	A	L	L

Table 5.1: TM domain sequences of low complexity scaffolds with additional motifs. The sequences are aligned to the heptad repeat schematic *a b c d e f g*, where red denotes the primary interface for the scaffold, and blue denotes the secondary binding interface.

5.2 Experimental methods

See chapter 3.

5.3 Results

5.3.1 Analysis of scaffold insertion and orientation

GALLEX chimera proteins were assessed for correct insertion using a combination of complementary tools. The first tool was the theoretical thermodynamic predictor, ΔGPredictor, which assigned an estimate of free energy of insertion through the Sec 61 translocon. The second tool makes use of high selectivity according to chimera orientation by growing *E. coli* NT326 transformed cells on minimal media. Finally, a sodium hydroxide wash (see section 3.7.1.4) was performed on the L17 scaffold and mutants in addition to both controls.

5.3.1.1 Theoretical insertion of scaffold TM domains

Typically, integral membrane proteins are accompanied by a chaperone to protect the hydrophobic residues from forming insoluble aggregates in the cytosol. The hydrophobic mismatch between

non-polar side-chains and the polar head group region of the lipid membrane acts as a thermodynamic barrier preventing spontaneous insertion (227). The integral membrane protein known as a translocon provides a mechanism to chaperone the TM protein and partition the membrane for TM protein insertion. The web based tool ΔGPredictor (66) was used to analyse the theoretical free energy of TM protein insertion into a biological membrane via the endoriticulum Sec 61 translocon. The predictive modelling tool ΔGPredictor is based on *in vitro* experimental results of the recognition of 19-residue long amino acid sequence through the Sec 61 translocon. The free energy difference is based on the assumption that individual amino acid contributions are additive, with the hydrophobic moment and length of the helix also taken into account. The predicted free energies of insertion (ΔG_{pred}) of each TM domain sequence are displayed in Figures 5.3, 5.4, and 5.5. In all cases the net ΔG_{pred} was negative suggesting that they would spontaneously insert across the bilayer. L17 was in complete favour of insertion due to the strong hydrophobic nature of leucine, with only the chimera protein linker region reporting a positive ΔG_{pred}. Adding a GG4 motif was particularly unfavourable yet the whole sequence was still theoretically prone to insert. Glycine plays a crucial role in TM helix packing, although it is also considered to be a helix breaker due to the flexibility the side chain provides the immediate peptide backbone (228). This would expose the hydrophilic peptide backbone to the nonpolar lipid alkyl tails. Replacing glycine with isoleucine was a considerably favourable substitution due to the difference in hydrophobicity. Finally, the addition of glutamine was predicted to be the least favourable residue to insert given its strong polar nature. Second to cysteine, glutamine is the least frequently occurring amino acid in α-helical TM domains (228).

The ΔG_{pred} of AZ2, as seen in Figure 5.3, predicts the favourable insertion of the TM peptide across the membrane. Even though alanine contributes to an increase in ΔG_{pred} due its borderline apolar nature (65), it is the second most common amino acid in TM α-helices and plays a significant role in helix-helix packing (28). Applying a GG4 motif makes AZ2 slightly less favourable, yet the combined hydrophobic contribution from the leucine residues still produces a negative net ΔG_{pred}. Mutation of either glycine residue to isoleucine encourages protein insertion by increasing the net hydrophobicity of the sequence. The inclusion of serine towards the centre of the TM domain increases the net ΔG of insertion yet the combined ΔG_{pred} still remains in favour of insertion. It is interesting to note that serine is the most frequently occurring non-charged polar residue in TM

α-helix composition (228). Serine may satisfy the need to hydrogen bond by forming an *intra*helical hydrogen bond with the backbone carbonyl oxygen atom (71).

AZ2(GG4)$_A$ has a ΔG_{pred} profile similar to AZ2(GG4)$_L$. The replacement of leucine to alanine in AZ2(GG4)$_A$L6A and AZ2(GG4)$_A$L10A demonstrates a decrease in the propensity for insertion given that leucine was predicted to contribute significantly to insertion due to its position on the hydrophobic scale.

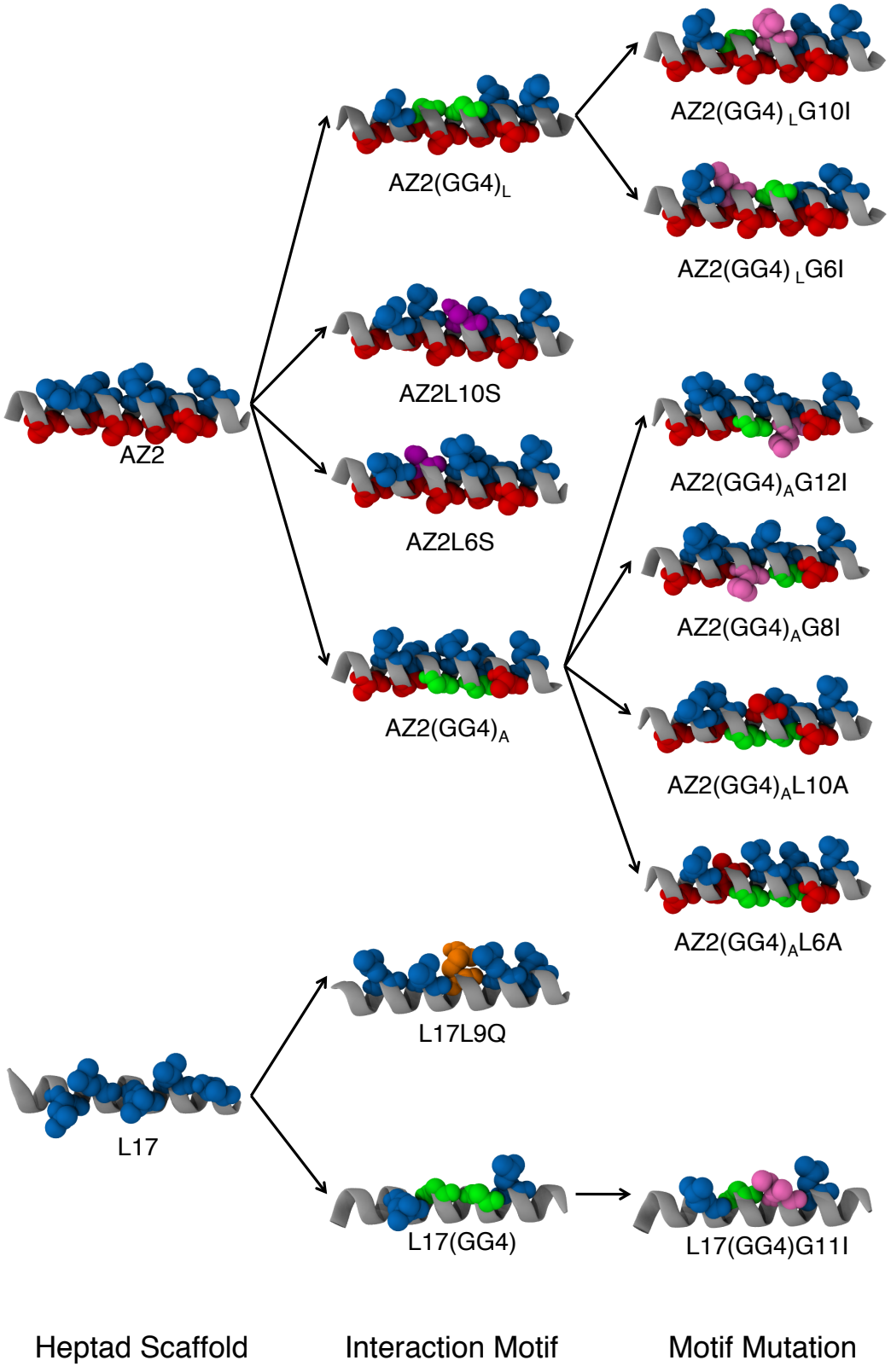

Figure 5.2: Atomistic representations of the single low complexity scaffolds used throughout this study. Residues at heptad repeat sides *a c d* and *f* are shown as space filling spheres. Residues are coloured according to their type: red - **alanine**, blue **leucine**, green - **glycine**, light purple - **isoleucine**, orange - **glutamin**, and dark purple **serine**.

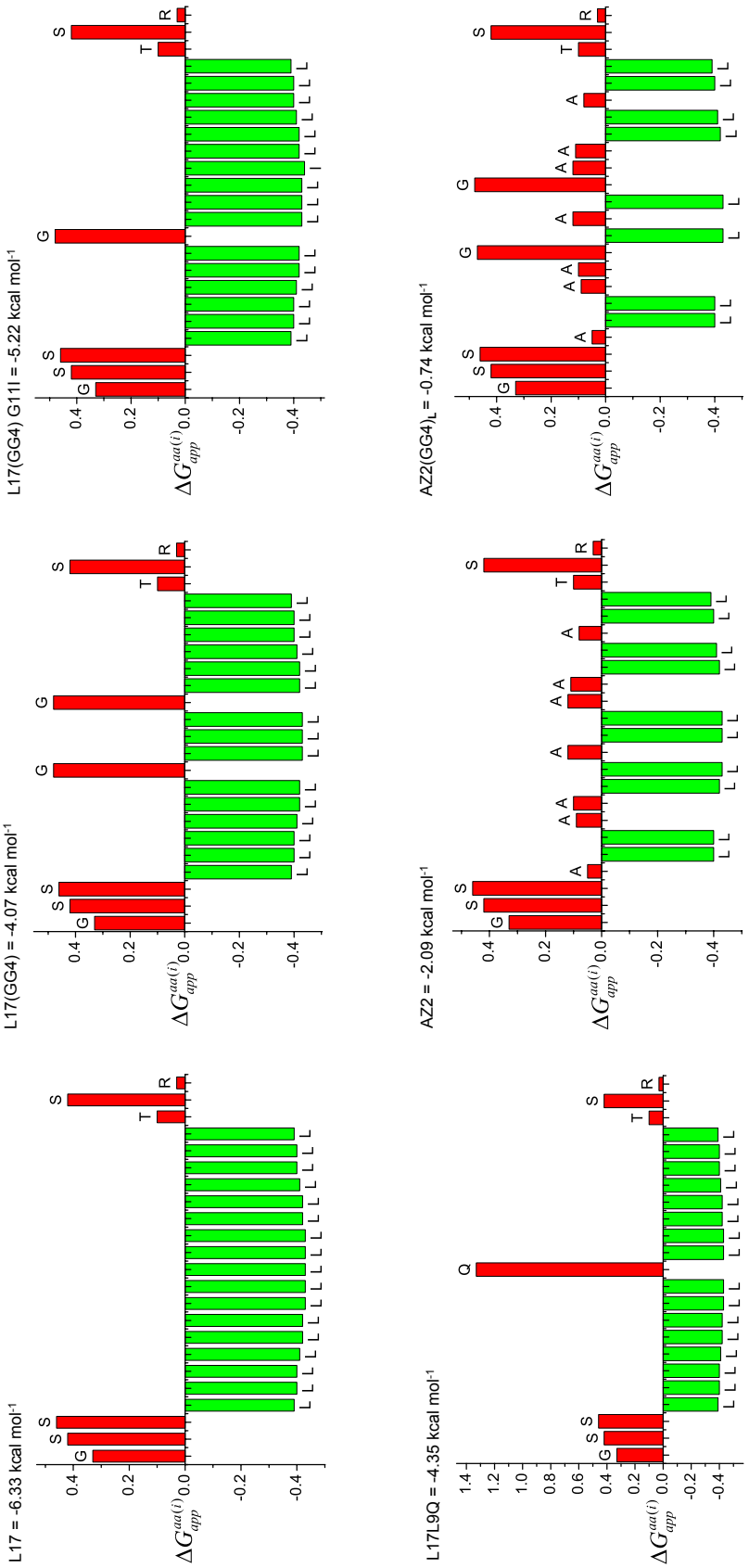

Figure 5.3: Theoretical ΔG_{pred} of TM domains L17, L17(GG4), L17(GG4)G11I, L17L9Q, AZ2, and AZ2$(GG4)_L$. Green indicates residues in favour of membrane insertion, and red indicates otherwise. For a brief explanation on theoretical ΔG_{pred}, see section 5.3.1.1 or Ref (66).

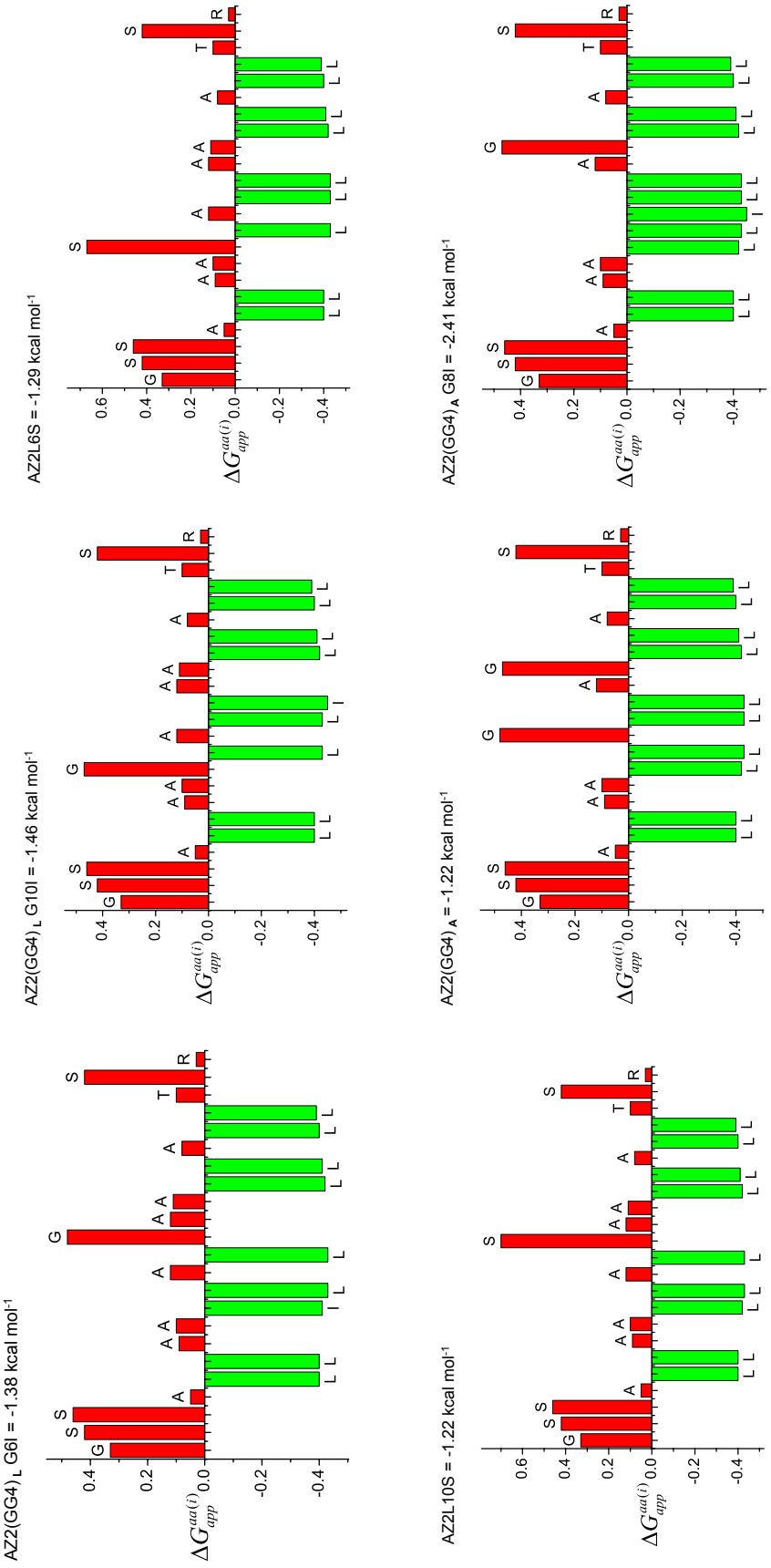

Figure 5.4: Theoretical ΔG_{pred} of TM domains AZ2(GG4)$_L$G6I, AZ2(GG4)$_L$G10I, AZ2L6S, AZ2L10S, AZ2(GG4)$_A$, and AZ2(GG4)$_A$G8I. Green indicates residues in favour of membrane insertion, and red indicates otherwise. For a brief explanation on theoretical ΔG_{pred}, see section 5.3.1.1 or Ref(64).

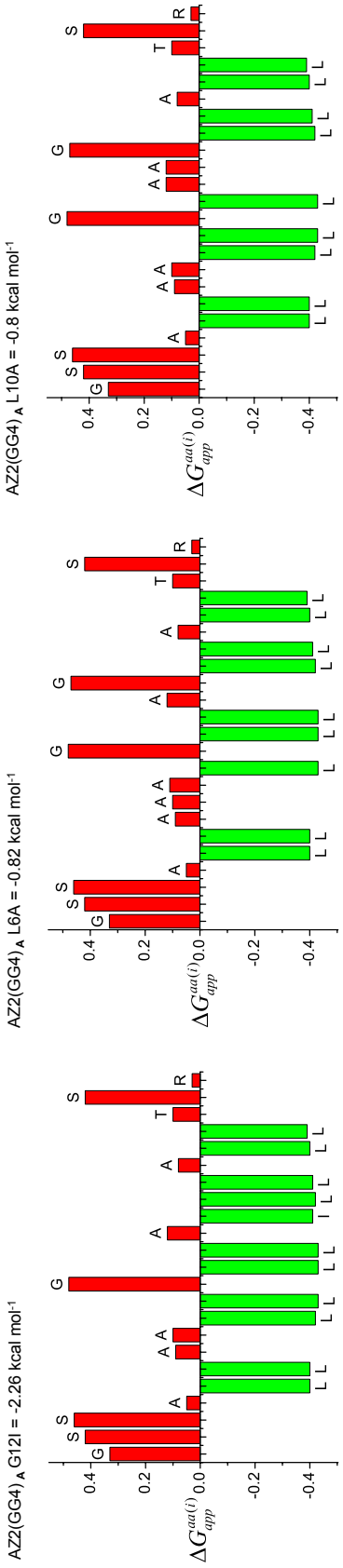

Figure 5.5: Theoretical ΔG_{pred} of TM domains AZ2(GG4)$_A$G12I, AZ2(GG4)$_A$L6A, and AZ2(GG4)$_A$L10A. Green indicates residues in favour of membrane insertion, and red indicates otherwise. For a brief explanation on theoretical ΔG_{pred}, see section 5.3.1.1 or Ref (66).

5.3.1.2 Analysis of protein chimera association with the inner membrane from sodium hydroxide washes

Attempts at quantifying the expression levels of L17 chimeric scaffolds in *E. coli* SU101 resulted in significant levels of protein over expression. Therefore, to ensure that the expressed L17 chimeric scaffolds and mutants were associating with the inner member and not forming cytosolic aggregates, cell cultures were washed with sodium hydroxide. Membranes washed with sodium hydroxide result in weakly associating proteins being removed, leaving strong membrane associating and integral membrane proteins behind.

Cell cultures were washed with ice-cold sodium hydroxide and centrifuged. If the chimeric protein associated with the *E. coli* inner membrane, it would be detected from western blotting in the whole cell (WC) fraction and membrane protein (MP) fractions with a close to undetectable amount of protein in the soluble protein (SP) fraction. Western blotting followed by immunoblotting with anti-MBP antibodies revealed that L17 chimera protein variants were present in the WC and MP fractions (Figure 5.6 (A)). As the MP fraction was very unclear, possibly due to protein over expression, a fresh membrane fraction was prepared and washed with acetate. After acetate precipitation and immunoblotting of three independent biological samples, all L17 chimera scaffolds were shown to be strongly associated with the inner membrane, as seen in Figure 5.6 (B).

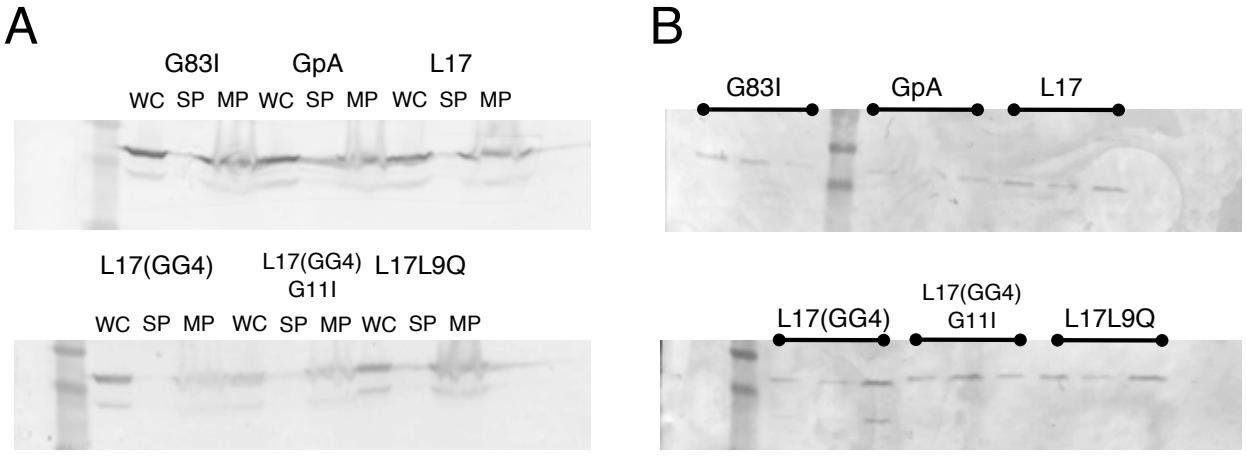

Figure 5.6: Immunoblots of polyleucine chimera scaffolds along with both controls G83I and GpA after (A) being washed with sodium hydroxide were loaded as whole cell fractions (WC), soluble proteins fractions (SP), and membrane associating protein fractions (MP). (B) Three independent samples of the MP were washed with sodium hydroxide then washed with acetate. All samples displayed membrane bound protein fractions after immunoblotting with anti-MBP antibodies.

5.3.1.3 MalE complementation assay

Prior to performing a β-galactosidase assay, it was imperative to establish whether the membrane associating chimeric proteins were orientated across the inner membrane with the MBP anchor in the periplasm and the LexA binding protein in the cytoplasm side. Chimeric proteins were expressed in MalE deficient *E. coli* NT326 by 10 μM IPTG induction and then grown on M9 minimal media containing 0.4% maltose. Cells expressing chimeric proteins incapable of crossing the inner membrane, or incorrectly inserted would show no sign of cell grown after approximately three days of incubation at 37°C. All chimeric proteins expressed in *E. coli* NT326 successfully grew on the minimal media plates as seen in (D) of Figures 5.7, 5.8, 5.9, and 5.10.

5.3.2 Analysis of low complexity scaffold interactions using GALLEX

The propensity for self-association of low complexity scaffolds and scaffolds with motif modifications were analysed using the GALLEX assay as described in section 3.8.1. The TM domain of GpA and the mutant G83I, were used as positive and negative controls, respectively. The seminal works by Russ *et al.,*(25), Schneider *et al.,*(26) and Brosig *et al.,*(24) reported similar difference in self-association between both controls in a biological context. The three basic low complexity scaffolds (L17, AZ2, AZ2$(GG4)_A$) and GpA control were normalised to G83I. Modifications of low complexity scaffolds were normalised to the value of their *native* sequence; for example, AZ2L6S was normalised to AZ2.

Each GALLEX signal was scaled to the WC protein expression level obtained using western blotting and immunoblotting. In order to remove technical error from every blot performed across this study, a complete blot of native MBP and chimera bound MBP was selected at random. The average band intensity of native MBP was calculated using ImageJ. The average intensity was used as a normalising factor for band intensities of MBP bound chimeric proteins across all blots. After normalising to the native MBP average intensity, a natural logarithm function was applied in order to remove excessive variation between protein expression values. Variation was typical of over development, change in scanner, and a difference in images resolution. An extensive survey on published western blots by Gassmann *et al.,* (229) highlight the issues associated with western blot protein quantification.

Low complexity scaffold chimeric proteins were subjected to three independent β-galactosidase measurements. The final result was taken from the average of the three samples and the error was given by the standard error of the mean. The individual measurements were tested against a gaussian distribution before being subjected to one-way ANOVA with a p-value of 0.05.

5.3.2.1 Low complexity scaffolds

The GALLEX signal for G83I was scaled to protein expression levels (a representation of three independent biological repeats loaded into a western blot, as seen in Figure 5.7 (B)) and normalised to 1 (Figure 5.7 (A)). The three low complexity scaffolds were scaled by protein expression levels and then normalised to the scaled value of G83I. The results are presented in Figure 5.7. GpA is approximately 30% of the G83I signal, within rough approximation of the original works by Schneider and Engelman (26). Polyleucine self-association was found to be slightly less favourable than G83I. The AZ2 scaffold was approximately 50% the signal of G83I, and according to ANOVA analysis in Figure 5.7 (C)), AZ2$(GG4)_A$ was statistically similar to GpA, suggesting a strong association.

5.3.2.2 Polyleucine scaffolds

GALLEX signals for the L17 scaffolds and three mutants: L17(GG4); L17(GG4)G11I; and L17L9Q, were scaled to protein expression levels from western blots (see Figure 5.8 (B) as a representation) and then normalised to L17 (Figure 5.8 (A)). One-way ANOVA analysis predicted a statistically insignificant difference between any of the sequences with the exception of L17L9Q (Figure 5.8 (C)).

5.3.2.3 Alanine zipper scaffolds

GALLEX signals for the AZ2 low complexity scaffold and mutants were scaled to protein expression levels (see Figure 5.9 (B) as a representation). The five mutants were normalised to the scaled value of AZ2 as depicted in Figure 5.9 (A). Introducing a GG4 motif on the leucine face of the helix (AZ2$(GG4)_L$) produced a significant drop in GALLEX signal, to approximately 80 % of AZ2. The isoleucine substitutions in the GG4 motif disrupted association resulting in a signal equivalent to AZ2. Replacing leucine with serine at 6c resulted in a significant drop in GALLEX

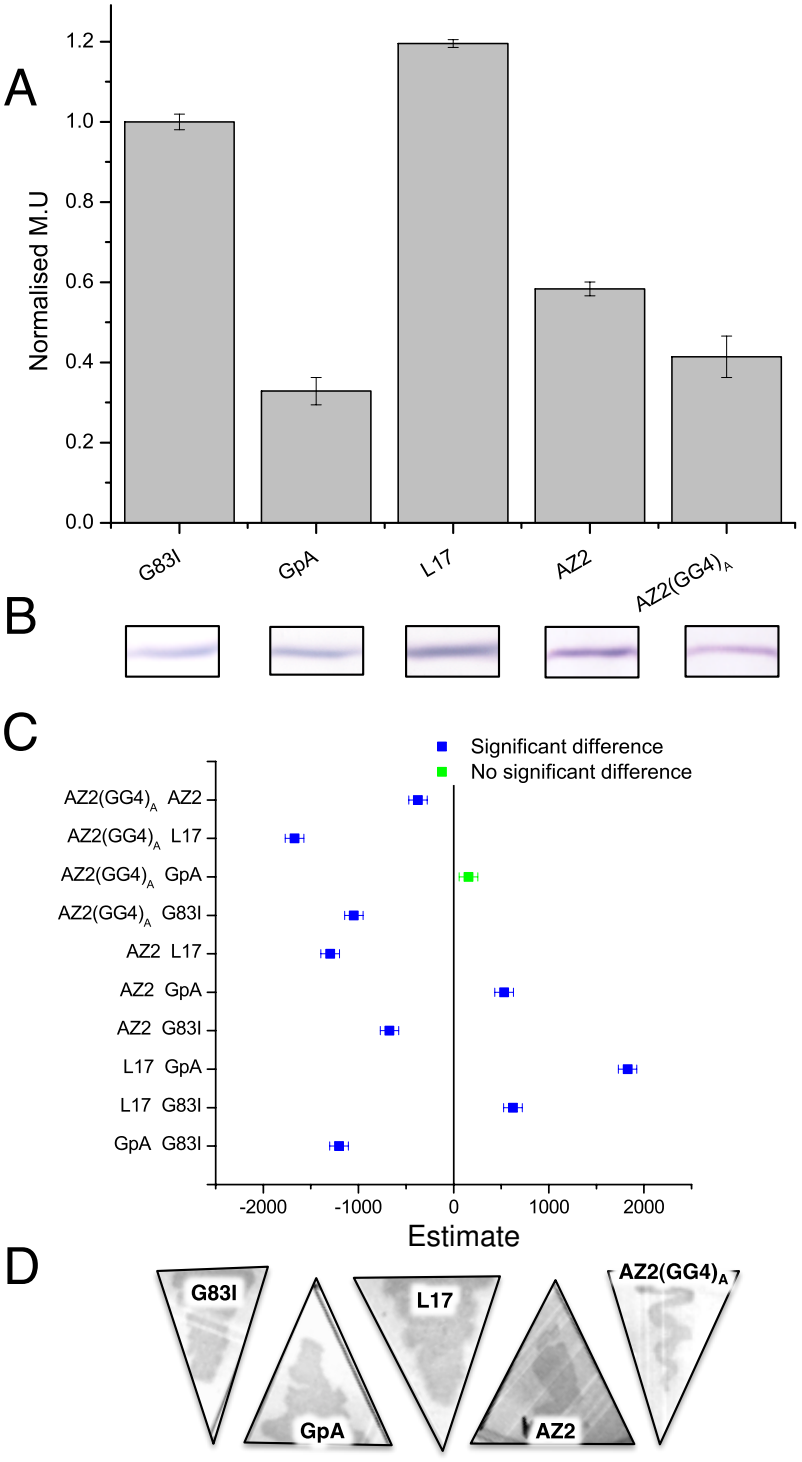

Figure 5.7: (A) Semi-quantitative GALLEX assay of low complexity scaffolds using sets of three independent biological repeats. (B) Averaged values were analysed by one-way ANOVA with a significance cut-off of 0.05. (C) GALLEX signals had been scaled by expression levels. The protein bands present are are a subset from the quantification of all biological repeats. (D) MalE complementation assays were performed to verify the orientation of integral membrane chimeric proteins.

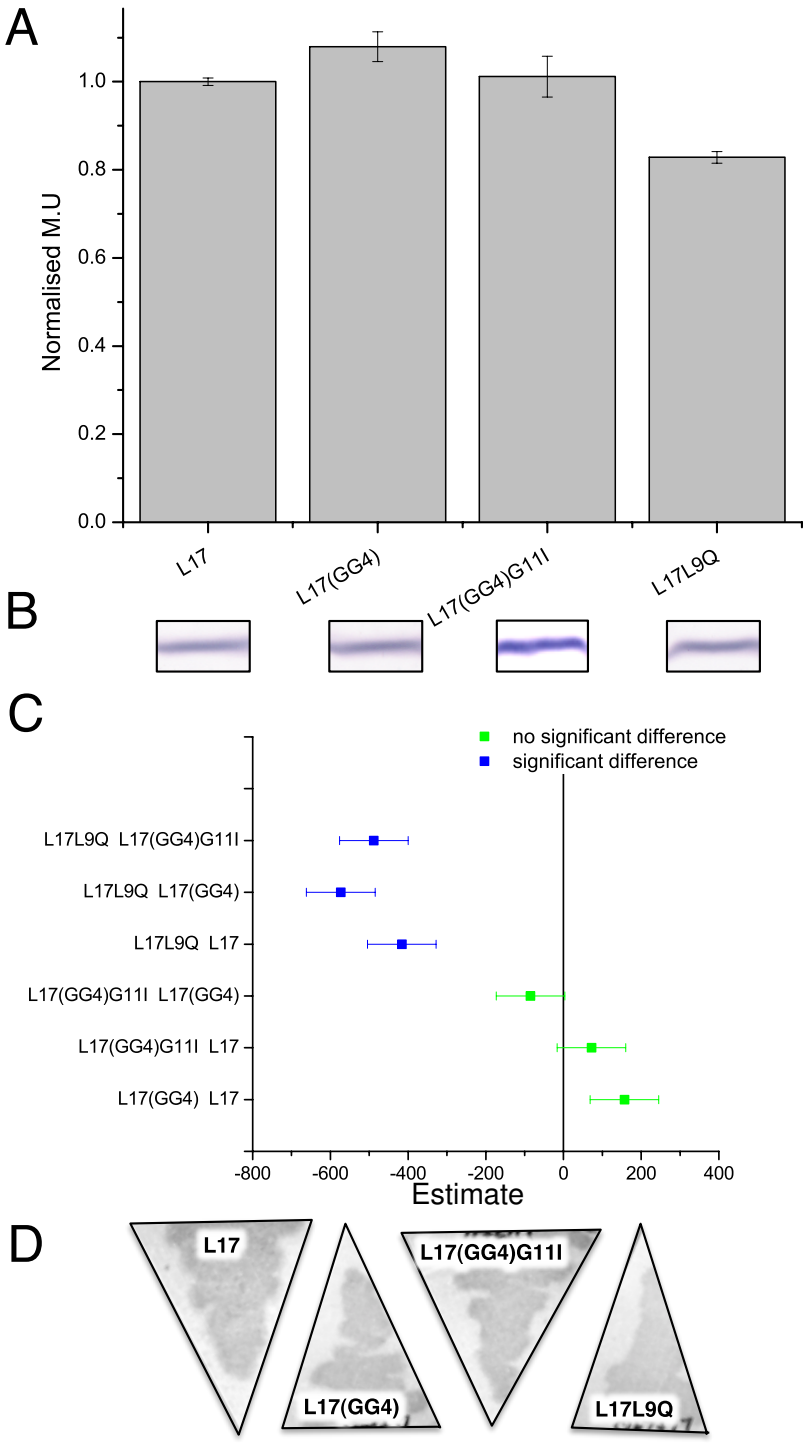

Figure 5.8: (A) Semi-quantitative GALLEX assay of L17 scaffolds using sets of three independent biological repeats. (B) Averaged values were analysed by one-way ANOVA with a significance cut-off of 0.05. (C) GALLEX signals had been scaled by expression levels. The protein bands present are are a subset from the quantification of all biological repeats. (D) MalE complementation assays were performed to verify the orientation of integral membrane chimeric proteins.

signal to approximately 80% of AZ2, similar to the propensity of self-association of AZ2(GG4)$_L$. Surprisingly, introducing serine further along the TM domain at position 10*g* resulted in drastic

reduction in GALLEX signal, approximately 35% of AZ2. In response to the changes made to AZ2 it is evident that association can be driven further by the inclusion of a GG4 motif, and even more so by substituting the sixth leucine for a serine.

5.3.2.4 Bifunctional alanine zipper scaffolds

The AZ2(GG4)$_A$ scaffolds and mutants were scaled to expression levels (see Figure 5.10 (B) as a representation) before being normalised to AZ2(GG4)$_A$ (Figure 5.10 (A)). Both variants of leucine to alanine substitutions were statistically similar to AZ2(GG4)$_A$ (Figure 5.10 (C)). On the other hand, substituting either glycines to isoleucines on the secondary interfaces resulted in a significantly large increase in GALLEX signal with no discernible difference between the two. In response to changes AZ2(GG4)$_A$, it is evident that substituting either glycines in the alanine associating and neighbouring GG4 significantly disrupts dimerisation. Conversely, replacing leucine for alanine makes no changes to the propensity for AZ2(GG4)$_A$ to form a dimer.

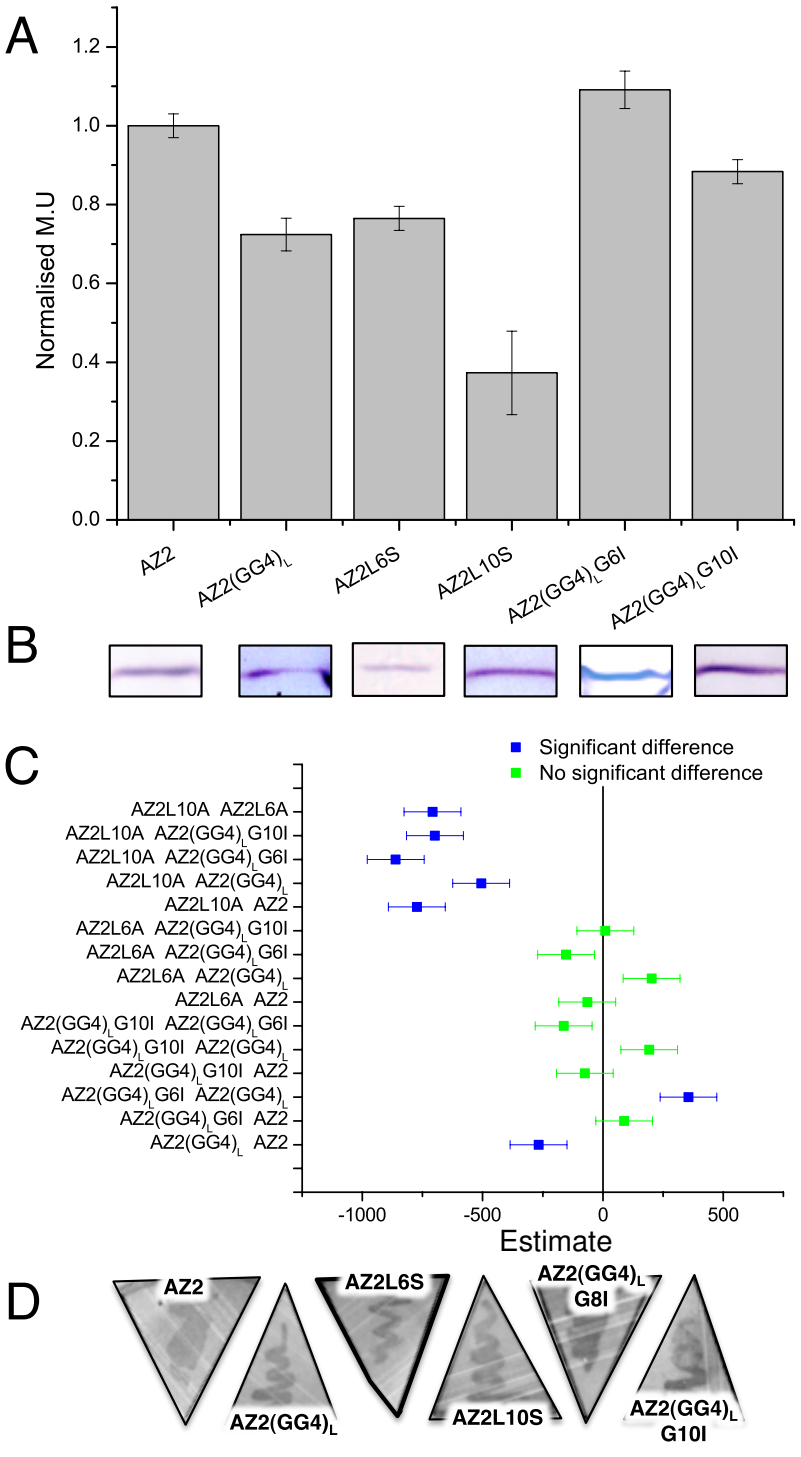

Figure 5.9: (A) Semi-quantitative GALLEX assay of AZ2 scaffolds using sets of three independent biological repeats. (B) Averaged values were analysed by one-way ANOVA with a significance cut-off of 0.05. (C) GALLEX signals had been scaled by expression levels. The protein bands present are are a subset from the quantification of all biological repeats. (D) MalE complementation assays were performed to verify the orientation of integral membrane chimeric proteins.

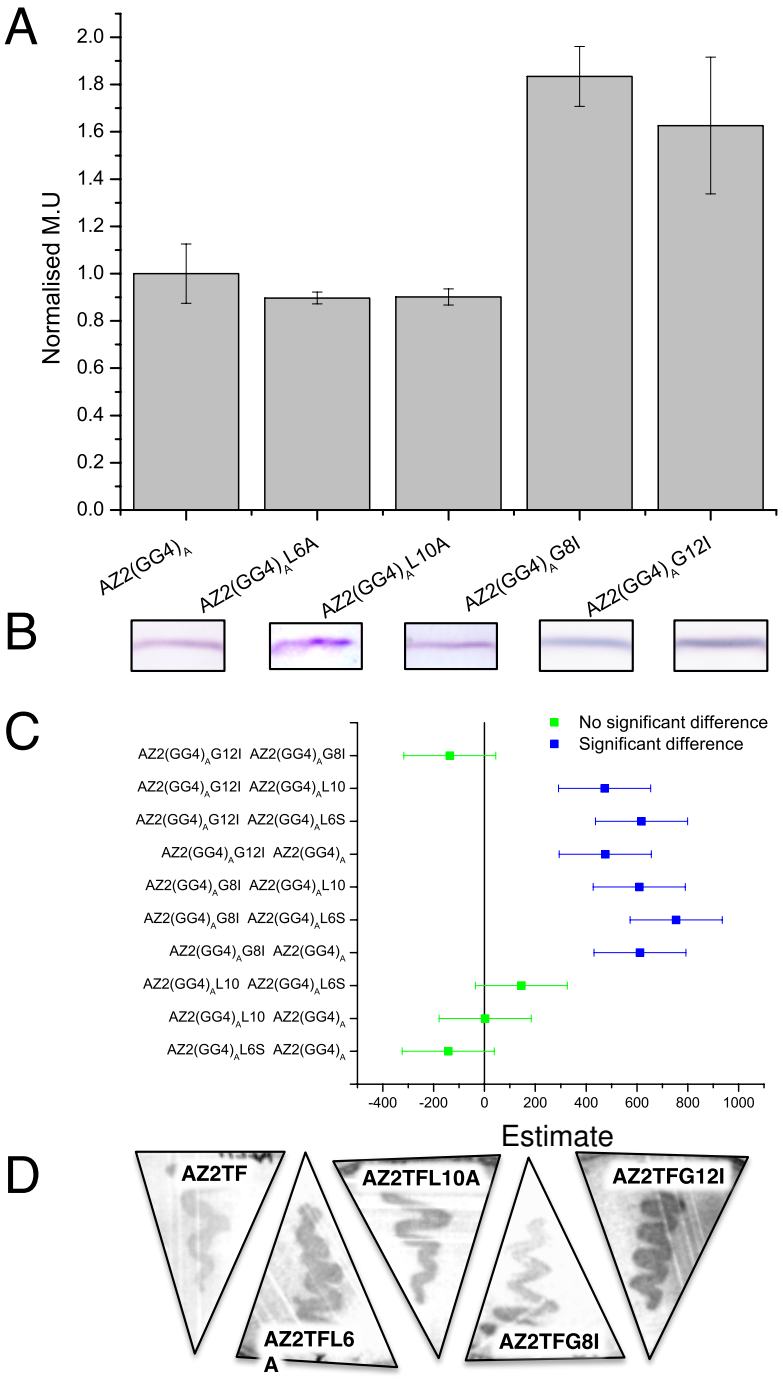

Figure 5.10: (A) Semi-quantitative GALLEX assay of AZ2(GG4)$_A$ scaffolds using sets of three independent biological repeats. (B) Averaged values were analysed by one-way ANOVA with a significance cut-off of 0.05. (C) GALLEX signals had been scaled by expression levels. The protein bands present are are a subset from the quantification of all biological repeats. (D) MalE complementation assays were performed to verify the orientation of integral membrane chimeric proteins.

5.3.3 Free energy of association in a biological membrane

Detailed knowledge of the equilibria and free energy characteristics of a *de novo* membrane protein in a biological context is essential. Even though there is a wealth of excellent biophysical reports on the free energy of TM domain association, the results are generally constrained by the simplistic mimetics used for solubilising membrane proteins. The composition and macrostructure of the mimetic may influence free energy results. Equally so, modelling the complex composition of a biological membrane using computational techniques is currently beyond the capacity of most high end clusters, high performance computers and current force fields.

This sections reports the relative Gibbs free energy calculations of self-association in a biological membrane. Both controls are measured, along with the: L17; L17(GG4); L17L9Q; AZ2; AZ2$(GG4)_L$; and AZ2$(GG4)_A$ constructs. Per construct, nine data points from three independent cell cultures were induced with 0.0001 mM - 1 mM IPTG. Cell cultures were subjected to western blotting followed by immunobloting using the anti-MBP antibody. A linear fit was applied to the averaged protein expression levels according to IPTG concentrations. Only samples from cell culture sets where the average protein expression approximated a linear fit would contribute to the IPTG concentration curve. The concentration curve is sigmoidal in appearance. The data was fit using a non-linear Hill function assuming a simple monomer-dimer equilibrium:

$$y = V_{max} \frac{x^n}{K_d^n + x^n}, \tag{5.1}$$

where K_d is the apparent dissociation constant (K_d^{app}). The Hill function approximates a logistic function when the IPTG concentration is logarithmically scaled. This is important when a range of concentrations result in saturation of the GALLEX signal over a few orders of magnitude.

In a biological system many parameters must be taken into consideration. The steepness of the curve will be the sum of several equillibra (non-specific interactions), such as the self-association between chimeric proteins and association with a whole host of native biological molecules. The system is complicated further by the increased surface density of the chimeric proteins. This not only affects the monomer-dimer equilibrium but also the absolute amount of active LexA. As there is no direct control over the other biologically relevant equilibria, the results are limited to a simple monomer-dimer model. In addition, all values are only relevant within the context of other chimeric

protein scaffold free energy calculations.

The K_D^{app} (μM), was calculated by rearranging the non-linear Hill fit. The apparent free energy of association, ΔG^{app}, was calculated using the relationship $\Delta G^{app} = -RTln(K_D^{app})$, where the temperature was set to 298 K and the ideal gas constant, R = 0.00918 kcal K^{-1} mol^{-1}.

Similar levels of protein expression were found between GpA, G83I, AZ2, $AZ2(GG4)_L$ and $AZ2(GG4)_A$, all of which followed a linear relationship with respect to IPTG concentration. The expression levels of L17, L17(GG4) and L17L9Q followed a linear relationship with respect to IPTG concentration but clearly show signs of over expression as discussed in section 5.3.1.2. Insert A of Figure 5.11 through to Figure 5.18 present the concentration curves for the (scaled to expression) GALLEX signal vs. IPTG concentration of the aforementioned chimeric proteins. At very low IPTG concentrations, AZ2 (Figure 5.16), AZ2(GG4) (Figure 5.17) and G83I (Figure 5.12) showed a great deal of statistical variation. Although each independent sample was inoculated with approximately the same volume of IPTG solution, the statistical variation may be due to only a small subset of cells being exposed to the very low concentrations of IPTG. Samples inoculated with 0 mM IPTG (ddH_2O only) have been excluded from the IPTG concentration curve due to the log scale. All concentration curves were trimmed of data points where the corresponding averaged protein concentration did not follow a linear fit.

By assuming a simple monomer-dimer equilibrium between GALLEX chimera proteins the K_D^{app} from the non-linear Hill fits of GpA (Figure 5.11) and G83I (Figure 5.11), were 5.42 μM and 21.7 μM, respectively. The $\Delta\Delta G^{app}$ between the controls differ by 0.82 kcal mol^{-1} in favour of GpA. Although this trend was in agreement with the earlier works of Finger *et al.,* (33), their reported K_D^{app} for GpA and G83I were 3.1 μM and 118 μM, respectively, a difference of 2.15 kcal mol^{-1}.

The K_D^{app} for the low complexity scaffold L17 (Figure 5.13) was 40.7 μM. Interestingly, after substituting two leucines for a GxxxG motif (Figure 5.14), the K_D^{app} increased to 64.5 μM. The $\Delta\Delta G^{app}$ of self-association from L17 to L17(GG4) was -0.27 kcal mol^{-1}. The free energy contribution from a single glutamine residue on an L17 scaffold showed a decrease of 18.9 μM in disassociation compared to L17, resulting in a $\Delta\Delta G^{app}$ of 0.37 kcal mol^{-1}. There was as little as -0.01 kcal mol^{-1} difference between L17L9Q and the G83I control.

The K_D^{app} for AZ2 was 16.11 μM, resulting in a ΔG^{app} of 2.44 kcal mol^{-1}. Compared with the L17

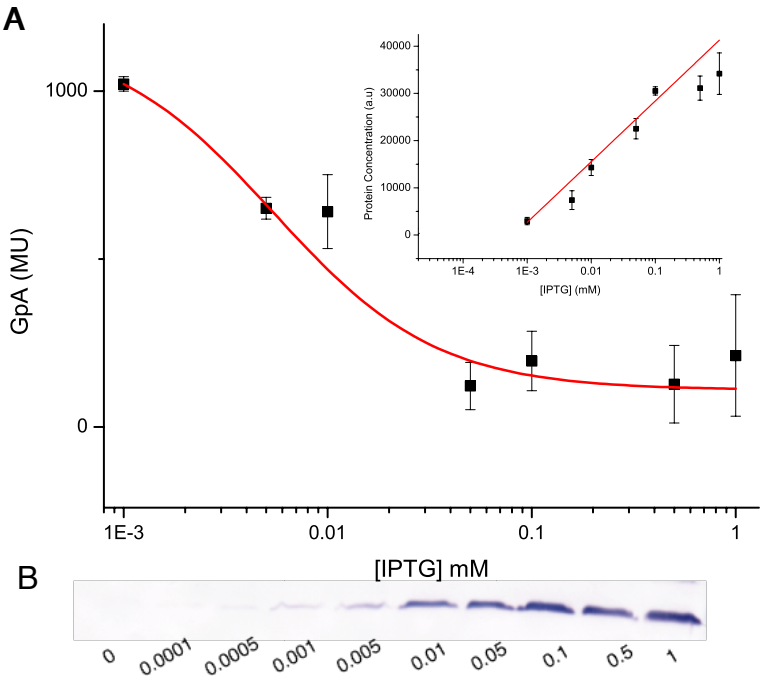

Figure 5.11: *(A)* Dimerisation of GpA transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). Cell cultures induced with 0.1 mM and 0.005 mM IPTG were removed. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

low sequence complexity scaffold, there was a $\Delta\Delta G^{app}$ of 0.55 kcal mol^{-1}. A GxxxG motif on the leucine face of AZ2 (Figure 5.17), drives self-association by 0.34 kcal mol^{-1}, 0.17 kcal mol^{-1} more than the basic AZ2 scaffold. This trend is in strong agreement with the standalone 10 μM IPTG induced GALLEX assay. AZ2(GG4)$_A$ drives self-association by 0.09 kcal mol^{-1} compared to AZ2 (Figure 5.18).

Construct	K_D^{app} (μM)	ΔG^{app} (kcal/mol)
G83I	21.7	6.34
GpA	5.42	7.15
L17	40.7	5.97
L17(GG4)	64.5	5.69
L17L9Q	21.8	6.33
AZ2	16.11	6.51
AZ2(GG4)$_L$	8.94	6.86
AZ2(GG4)$_A$	13.9	6.60

Table 5.2: Self-association free energy of low complexity scaffolds in a biological membrane. The K_D^{app} is an apparent disassociation constant from a simplified monomer-dimer equilibrium approximation.

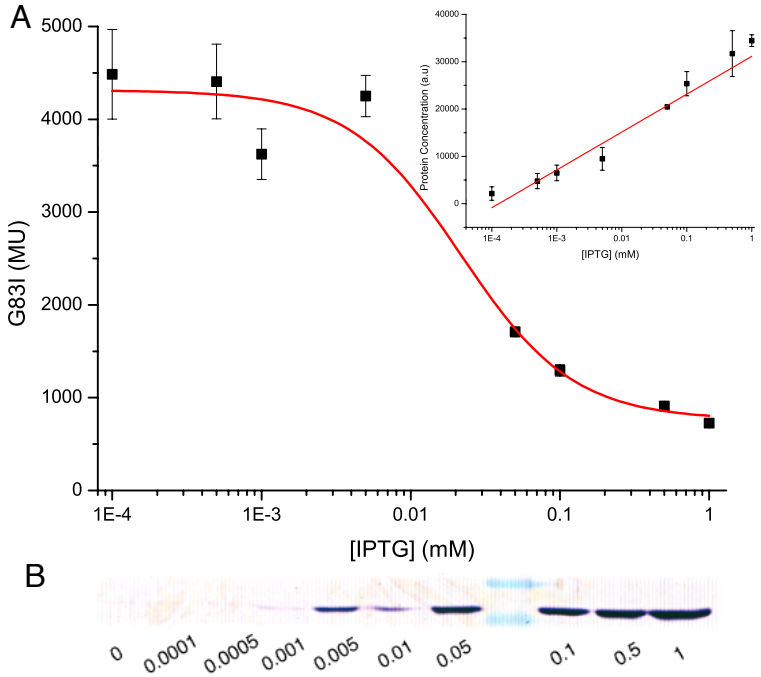

Figure 5.12: *(A)* Dimerisation of G83I transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). Cell cultures induced with 0.01 mM and 0.005 mM IPTG were removed. The error bars at 1 mM, 0.5 mM, 0.1 mM and 0.05 mM were small enough to exclude for clarity. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

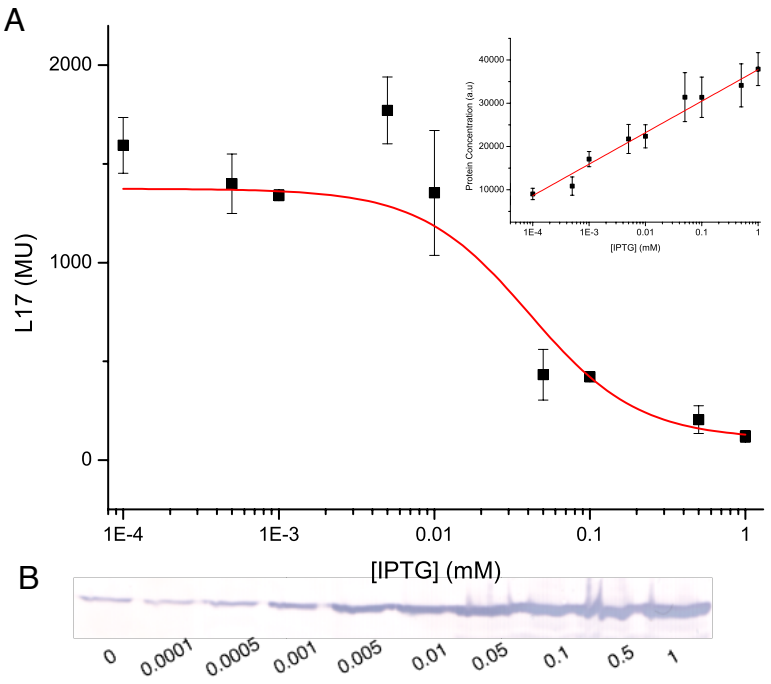

Figure 5.13: *(A)* Dimerisation of L17 transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). The error bar at 1 mM were small enough to exclude for clarity. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

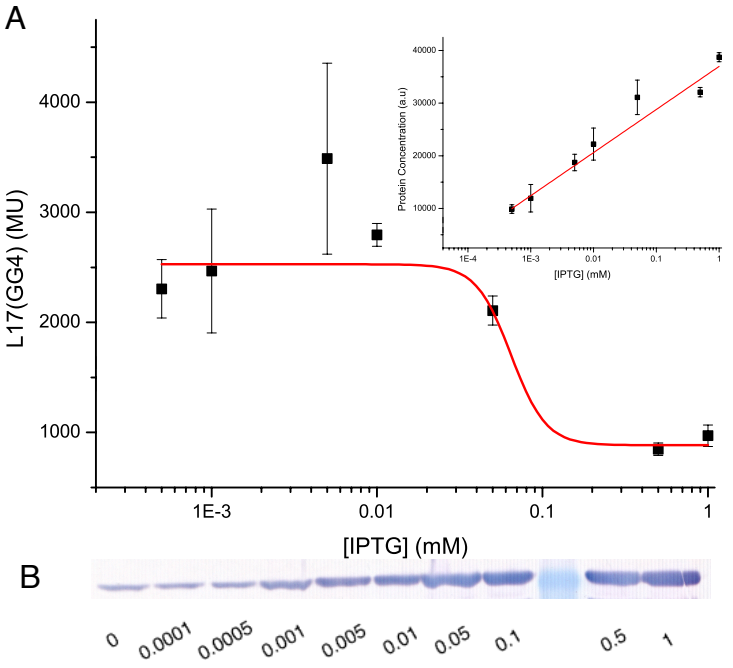

Figure 5.14: *(A)* Dimerisation of L17(GG4) transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). Cell cultures induced with 0.1 mM and 0.001 mM IPTG were removed. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

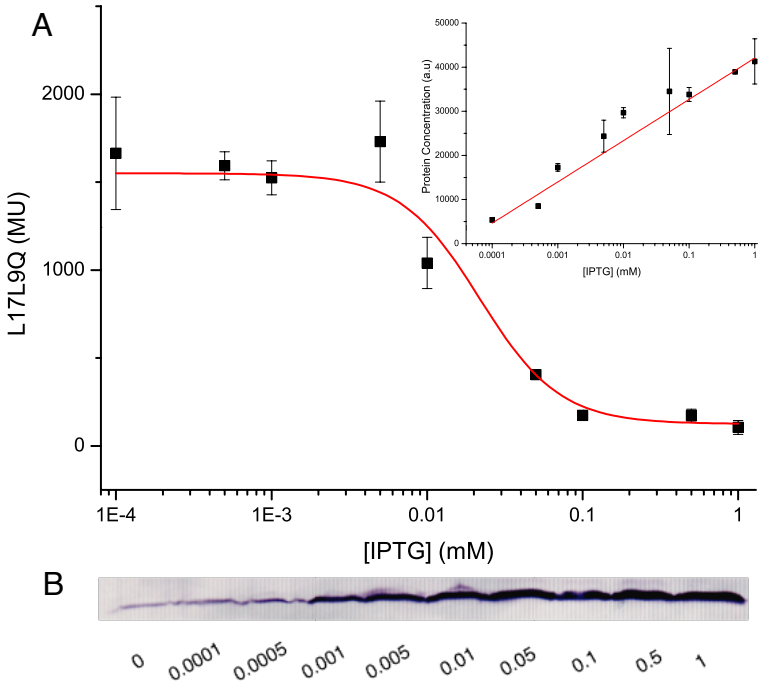

Figure 5.15: *(A)* Dimerisation of L17L9Q transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). The error bar at 1 0.5 mM, 0.1 mM, and 0.05 mM were small enough to exclude for clarity. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

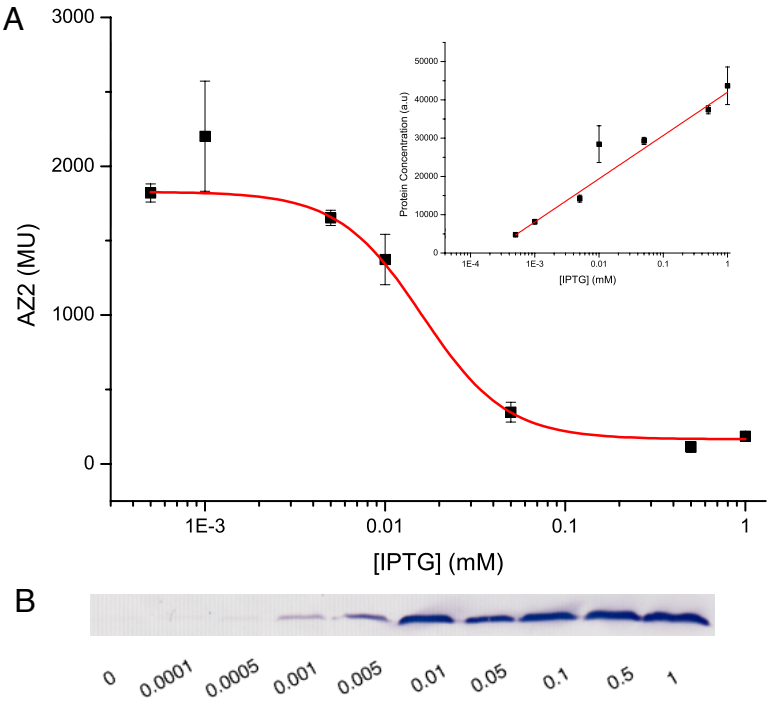

Figure 5.16: *(A)* Dimerisation of AZ2 transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). Cell cultures induced with 0.1 mM and 0.05 mM IPTG were removed. The error bar at 1 mM and 0.5 mM were small enough to exclude for clarity. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

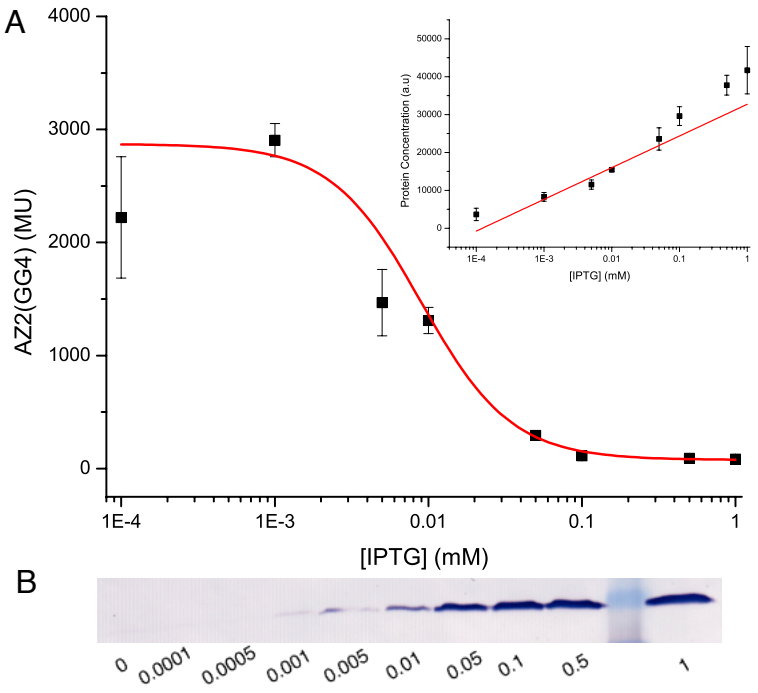

Figure 5.17: *(A)* Dimerisation of AZ2(GG4)$_L$ transmembrane helix in a biological membrane. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). Cell culture induced with 0.0005 mM were removed. The error bar at 1 mM, 0.5 mM, 0.1 mM, and 0.05 mM were small enough to exclude for clarity. *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

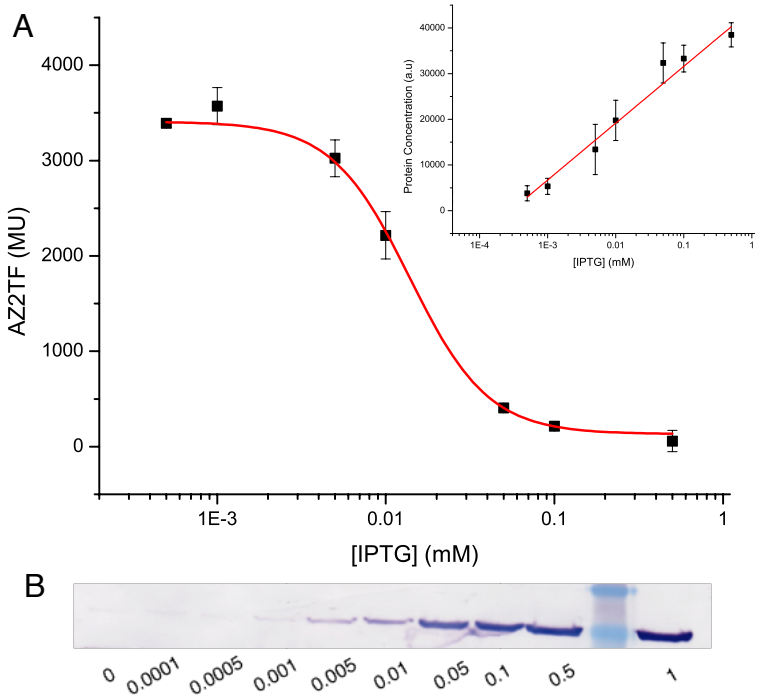

Figure 5.18: *(A)* Dimerisation of AZ2(GG4)$_A$ transmembrane helix in a biological membrane. Cell culture induced with 0.5 mM abd 0.0001 were removed. The error bar at 0.1 mM, and 0.05 mM were small enough to exclude for clarity. β-galactosidase activity (GALLEX signal) was measured to IPTG concentration (log scale). *(B)* Relative protein expression was resolved by immunoblotting western blots using anti-MBP and then quantified through densitometric analysis.

5.4 Discussion

Semi-quantitative β-galactosidase measurements from a 10 μM IPTG induced GALLEX assay were used to measure the propensity for self-association of low complexity scaffolds. Well-known helix-helix interaction motifs were then incorporated into these low sequence complexity scaffolds to determine the contribution a motif makes during self-association. Single point mutations were then made to the interaction motifs to determine whether an alternative binding interface existed to compensate for disruption to the primary binding helix-helix interface. Relative free energy values in a biological membrane were then calculated by extending the GALLEX assay to span a range of IPTG concentrations.

This discussion looks at the data presented in this chapter from two different perspectives. The first, how the low complexity scaffolds and inclusion of interaction motifs compare with relevant results available in the literature. The second, how the semi-quantitative GALLEX results and apparent relative free energy calculations can be used to generate a scaffold-motif map (see Figure 5.19). If the inclusion of a motif encouraged self-association according to the 10 μM IPTG induced standalone GALLEX assay then the result is represented by a green arrow. Changes, which demonstrate a decrease in self-association is represented as a red arrow, and any change which is statistically insignificant, is represented as a black arrow.

5.4.1 Polyleucine scaffolds

One-way ANOVA analysis of the low complexity scaffold L17 compared with the G83I negative control suggests that statistically L17 forms a weaker oligomer. In addition, $\Delta\Delta G_{app}$ of L17 is 0.38 kcal mol^{-1} in favour of G83I. Zhous *et al.,* (230) used SDS-PAGE analysis to show that a twenty-three long polyleucine peptide runs as a monomer in a detergent micelle. This was complemented by TOXCAT analysis of a chimeric protein with a thirteen long polyleucine TM domain. The signal was approximately 10% to 20% of a GpA control, a significantly lower signal than the L17 to GpA signals reported in this study. On the other hand, the work of Herrmann *et al.,* (79) reported that a sixteen residue polyleucine TM domain of the TOXR chimera protein gave a signal approximately 40% to 50% of GpA. Although this is marginally stronger than L17 to the GpA control in this study, both results could be classified as poor yet interacting dimers. Finally, further SDS-PAGE analysis

of a twenty residue long polyleucine TM peptide by Ruan *et al.,* (98) clearly indicates a monomer. Given the similarities between the results recorded in this study and those reported in literature, a polyleucine low complexity scaffold is considered to be predominately monomeric.

The incorporation of a GG4 interaction motif to the polyleucine low complexity scaffold was the first of three L17 modifications. One-way ANOVA analysis indicate statistically similar GALLEX signals between L17 and L17(GG4), yet the $\Delta\Delta G_{app}$ is 0.27 kcal mol^{-1} weaker than the basic polyleucine scaffold. SDS-PAGE analysis by Orzaez *et al.,* (51) suggest that a seventeen residue long polyleucine scaffold with a GG4 motif at the centre runs as a monomer. Their results compare well with the results presented in this study. It is evident that GG4 does not promote self-association on a polyleucine low complexity scaffold given the statistically equivalent signals between L17(GG4) and the second modification, the single-point modification G11I. Double alanine mutations to the TM domain of GpA by Doura and Fleming (48), found that GG4 mediated dimerisation was sequence context dependent. Substituting G_{79} and G_{83} with leucine did not sufficiently reduce dimerisation. Yet, the simultaneous mutations, L_{75}A and T_{87}A, result in GpA monomers. Therefore, without long-range interactions (those over 12 Å), it was suggested that alone, GG4 is not sufficient for dimerisation. In addition, TOXR activity assays of the GpA TM domain, found that L_{75}, I_{76}, V_{80}, V_{84}, and T_{87} were required to assist GG4 mediated oligomerisation (24). The inclusion and subsequent mutation of a GG4 in association with a polyleucine low complexity scaffold is depicted in Figure 5.19 as two black arrows branching from L17.

The propensity for self-association maybe affected depending upon where a polar residue occurs in a sequence (75). In the context of the polyleucine low complexity scaffold, helix-helix binding is assumed to be non-specific. If a single polar residue is not affected by the native van der Waals packing then it should be at the helix-helix interface, shielded from the low dielectric unfavourable bilayer core. According to Dawson *et al.,* (78), TOXCAT analysis demonstrated that the substitution of a single glutamine in the tumour necrosis factor 5 (ligand) was enough to drastically disrupt dimerisation. The third motif of the polyleucine low complexity scaffold in this study incorporates the polar residue glutamine at its centre. The GALLEX analysis presented here indicate a marginal, yet statistical significant increase in self-association compared with L17, with 0.37 kcal mol^{-1} in favour of the glutamine substitution; however, compared with the G83I negative control, there is almost no difference. Therefore, glutamine on a low complexity scaffold is not enough to show

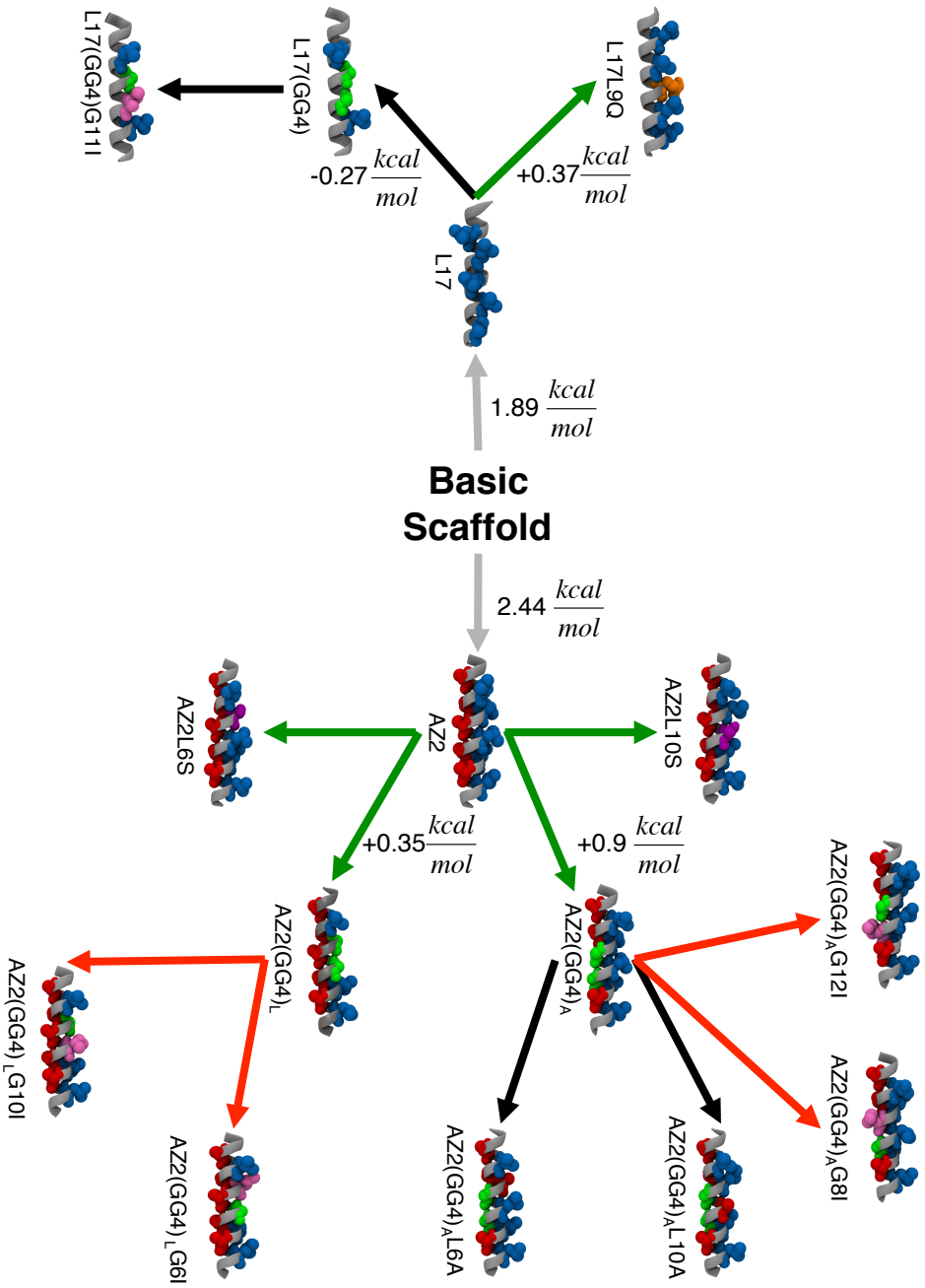

Figure 5.19: A thermodynamic 'road map', depicting how GG4 and polar residues contribute to the stability of two low complexity scaffolds. The changes in free energy between structures are derived from those calculated in this chapter. The colour of the arrows indicate how the motif contributes to self-association from the standalone 10 μM GALLEX results. A green arrow suggests a favourable contribution to self association whilst a red arrow suggests a destabilising contribution. A black arrow suggests no change.

strong associations between TM domains. The two most obvious reasons are: side-chain/side-chain hydrogen bonds competing with alternative hydrogen bonds between side-chain and backbone (231); or, the packing of neighbouring leucine side-chains prevents an ideal angle and distance between the two glutamine side-chains (232, 233). Once the hydrogen bond moves beyond 2 Å it

becomes a weak electrostatic interaction (78). With the supporting evidence from literature in mind, a tentative positive enforcement to self-association is considered based purely on the $\Delta\Delta G_{app}$ with L17. This is depicted as a green arrow branching from L17 in Figure 5.19.

5.4.2 Alanine zipper scaffolds

The second low complexity scaffold in this study was the alanine zipper. One-way ANOVA analysis of AZ2 according to the results of the 10 μM IPTG induced GALLEX signals clearly indicate that AZ2 is a stronger oligomer than either the G83I negative control or the L17 low complexity scaffold by as much as 0.17 kcal mol^{-1} and 0.55 kcal mol^{-1}, respectively. The results presented in this study coincide with reports from a POSSYCAT assay by Ridder *et al.,* (104). They demonstrated that their alanine zipper gave a signal twice that of a sixteen residue long polyleucine, and a signal 40 % of GpA.

The suggestion that a single serine residue was incapable of promoting significant helix-helix association was proposed by Gray and Matthews (234) and Ballesteros *et al.,* (235). A single serine would form an *intra*helical hydrogen bond with the $i-3$ or $i-4$ carbonyl oxygen atom. The formation of *inter*helical hydrogen bonds using a single serine would require the breaking of a backbone hydrogen bond. After the TOXCAT analysis of a randomised library of TM domains, Dawson *et al.,* (71) suggested that the frequently occurring motifs SxxxSSxxT and SxxSSxxT were the minimum requirements for any significant serine-based helix-helix association. The results presented in this chapter based on two single point substitutions in an alanine zipper low complexity background; L6S and L10S, suggests that serine stabilises self-association of an alanine zipper. One-way ANOVA analysis of both serine substitutions results in a statistically significant drop in GALLEX signal compared with the G83I control and native AZ2. Further still, the GALLEX signal strength of AZ2L10S is comparable with the GALLEX signal strength of GpA. The difference in the results presented here and the reports in literature may be the result of a more sophisticated mechanism behind the involvement of serine in TM domain interactions. For example, the involvement of serine in helix-helix interactions maybe sequence dependent. Based purely on the GALLEX results, the inclusion of a serine residue onto an alanine zipper low complexity scaffold is considered to stabilise helix-helix dimersation and is therefore depicted by a green arrow from AZ2 in Figure 5.19.

The arrangement of polar residues at strategic sites of a heptad repeats sequence has been well established (236, 237, 79, 69, 70, 75). However, the incorporation of a GG4 motif in the context of an alanine zipper heptad repeat low complexity scaffold is less well established, more so in the case of free energy calculations in a biological system.

Statistical analysis and TOXCAT analysis indicate that GG4 is one of the most over-represented pair of residues and is a major *inter*helical packing motif (28). Most dimers containing a pair of GG4 motifs adopt a right-handed helix crossing angle (238, 239). Senes *et al.,* (240) identified parallel and anti-parallel left-handed coiled-coils stabilised by the packing of a GG4 motif. Left-handed coiled-coils containing highly conserved glycine residues occur in the class II MHC α- and β-chains as identified by Cosson and Bonifacino (241).

According to the 10 μM IPTG induced GALLEX results and the biological free energy calculations, AZ2(GG4)$_{\mathrm{L}}$ is by 0.35 kcal mol^{-1} in favour of self-associating compared with AZ2. The statistical difference in GALLEX signals between a single glycine to isoluecine mutation of AZ2(GG4)$_{\mathrm{L}}$ reveals that a GG4 motif contributes to dimeric stability. Substitution of two alanine residues with a GG4 motif (AZ2(GG4)$_{\mathrm{A}}$) was shown to significantly drive self-association according to 10 μM IPTG induced GALLEX assays. Free energy calculations was 0.9 kcal mol^{-1} in favour of AZ2(GG4)$_{\mathrm{A}}$ compared with the AZ2 low complexity scaffold and 0.55 kcal mol^{-1} compared with AZ2(GG4)$_{\mathrm{L}}$. A green arrow from AZ2 in Figure 5.19 depicts both GG4 substitutions to the leucine and alanine face of AZ2.

Residues with constrained side-chain orientations, such as the β-branched alanine, valine and isoluecine are statistically more prevalent as neighbours to the GG4 motif (242). Although leucine constitutes just under a quarter of the average compositions of a TM α-helix, being able to sample more conformations, the γ-branched side-chain plays only a secondary role (28). Given that AZ2 adopts a left-handed helix crossing, and GG4 motifs are usually found in right-handed TM dimers, the slight increase in helix-helix propensity between AZ2 and AZ2(GG4)$_{\mathrm{L}}$ maybe indicative of steric interference between glycines due to the entropic penalty of neighbouring leucine side-chains. On the other hand, alanine side-chains (low on rotomeric entropic contributions) neighbouring a GG4 motif (AZ2(GG4)$_{\mathrm{A}}$) may satisfy a right-handed helix cross maximising glycine van der Waals packing.

5.5 Conclusion

The standalone GALLEX results and calculation of apparent free energy in a biological system are the first of their kind for these sequences (excluding the two controls). Although this is not the first time a GG4 motif has been studied in the context of a coiled-coil, it is the first time to the knowledge of the author that two different positions of the GG4 helix on an AZ2 low complexity scaffold have been compared. These experiments suggest that very few single changes to a basic polyleucine zipper can be made to encourage self-association. In addition, in the presence of a GG4, the alanine zipper can be stabilised further by placing the motif in association with the alanine face and that a single serine residue can encourage dimerisation. Finally, these results have been used to begin a rational design schematic of changes to oligomerisation upon the modification of a low complexity scaffolds. In the following chapter, the packing of side-chains and helix-to-helix orientations are examined in order to complement and help explain the rational design.

Chapter 6

Free energy of association of TM peptides in a POPC bilayer: comparison of coarse-grained and united-atom force fields

6.1 Introduction

Biological membranes are complex, dynamic, and fundamental to cell physiology. The cell is compartmentalised to maintain pH, and to permit selective transport of matter and energy between compartments and the outside environment via a sophisticated array of transport mechanisms (243). Local lipid arrangement is crucial to membrane-mediated protein-protein interactions (244, 245), and is typically driven by a mismatch between the side chain polarity and the nonpolar lipid environment. Differences between the hydrophobic length of the hydrophobic region of the TM protein domain and the hydrophobic tail region of the bilayer is defined as: a negative mismatch, where the hydrophobic span of the TM domain is shorter than the hydrophobic core of the bilayer; and a positive mismatch, where the hydrophobic span is longer than the hydrophobic core of the bilayer. In either case, the unfavourable contacts between TM domain and lipid can be satisfied by rearrangement of the annular and nearby lipids or an adjustment to the TM domain itself. The

negative mismatch, as illustrated in Figure 6.1, can be resolved by: (i) expelling part of or the complete protein (178); (ii) adopting a nonlameller arrangement around the TM domain (246); (iii) ordering of the alcyl tails of annular and nearby lipids (247, 248); and (iv) uncoiling or rearranging TM domain interactions (247). There are also reports of polar side chains snorkelling (93, 247, 246). The positive mismatch, as illustrated in Figure 6.2, can be resolved by: (i) expelling part of the TM domain or the complete TM domain (247); (ii) tilting the TM domain to match polarity between side chains and lipids (247, 249); (iii) ordering of the alcyl tails of annular and nearby lipids (247); and (iv) adjusting the helical turn (for example, from an α-helix to a π-helix) or rearranging *inter*helical TM domain interactions (247).

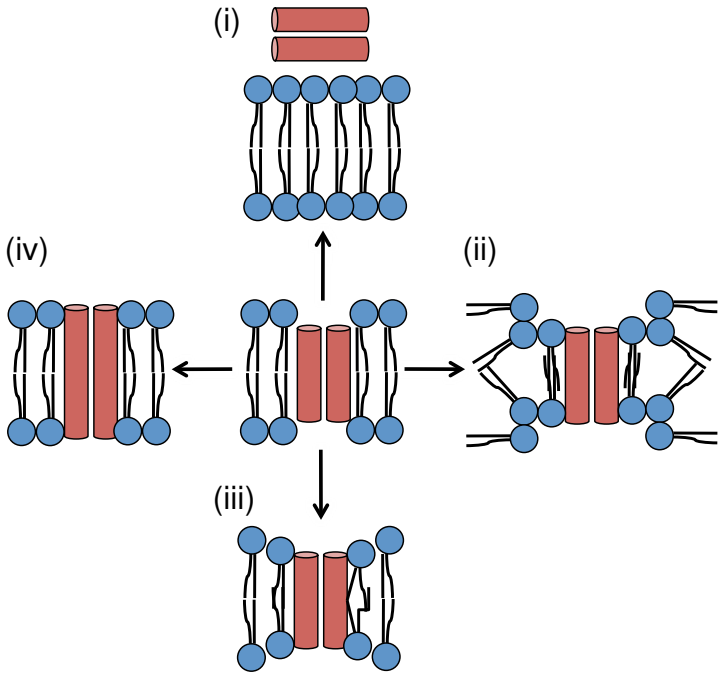

Figure 6.1: Negative hydrophobic mismatch between a bilayer and a dimer with a short hydrophobic stretch. The bilayer and dimer can correct the mismatch in a number of ways: (i) expelling the dimer from the bilayer; (ii) forming a nonlameller lipid environment; (iii) ordering the lipids along the hydrophobic stretch of the dimer; and (iv) rearranging the dimer interactions. Image adapted from Kandasamy *et al.,* (247).

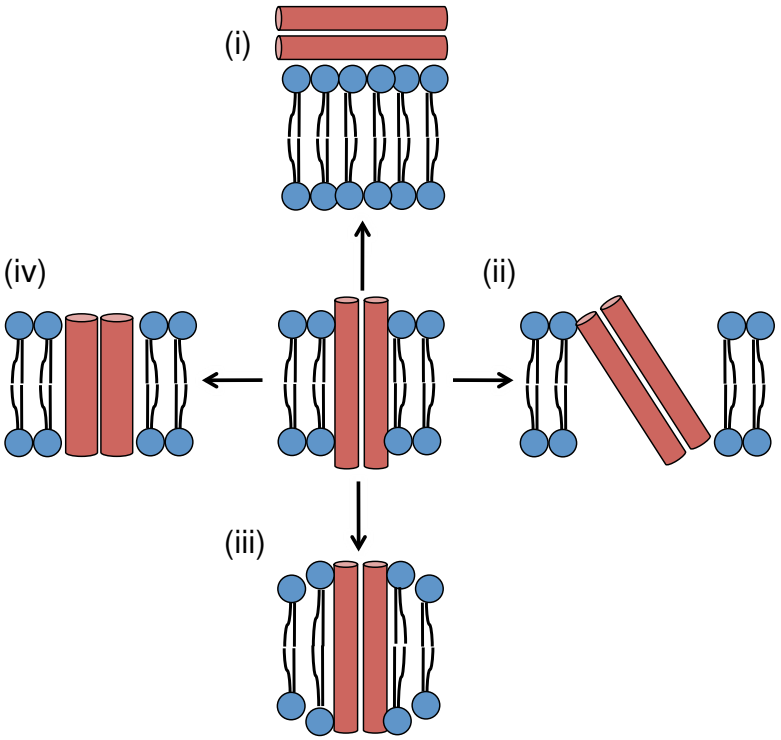

Figure 6.2: Positive hydrophobic mismatch between a bilayer and a dimer with a long hydrophobic stretch. The bilayer and dimer can correct the mismatch in a number of ways: (i) expelling the dimer from the bilayer; (ii) tilting the dimer; (iii) ordering the lipids along the hydrophobic stretch of the dimer; and (iv) rearranging the dimer interactions. Image adapted from Kandasamy *et al.,* (247).

In chapter 4, the free energy profile of two associating α-helical TM peptides demonstrated how the ordering of annular and nearby lipids affected the convergence of the PMF profile. The original publication of the Gromacs WHAM (198) code clearly illustrates a set of PMF profiles normalised to 0 kJ mol^{-1} at the point at which each profile plateaus. This is a fundamental step in calculating the difference in free energy between two systems. Both PMF profiles must be normalised to a reference point where both profiles do not change (as seen in Figure 6.3 (C)). Failure to do so, whether due to limitations of the system box and subsequent reaction coordinate, can result in a biased comparison between free energy values.

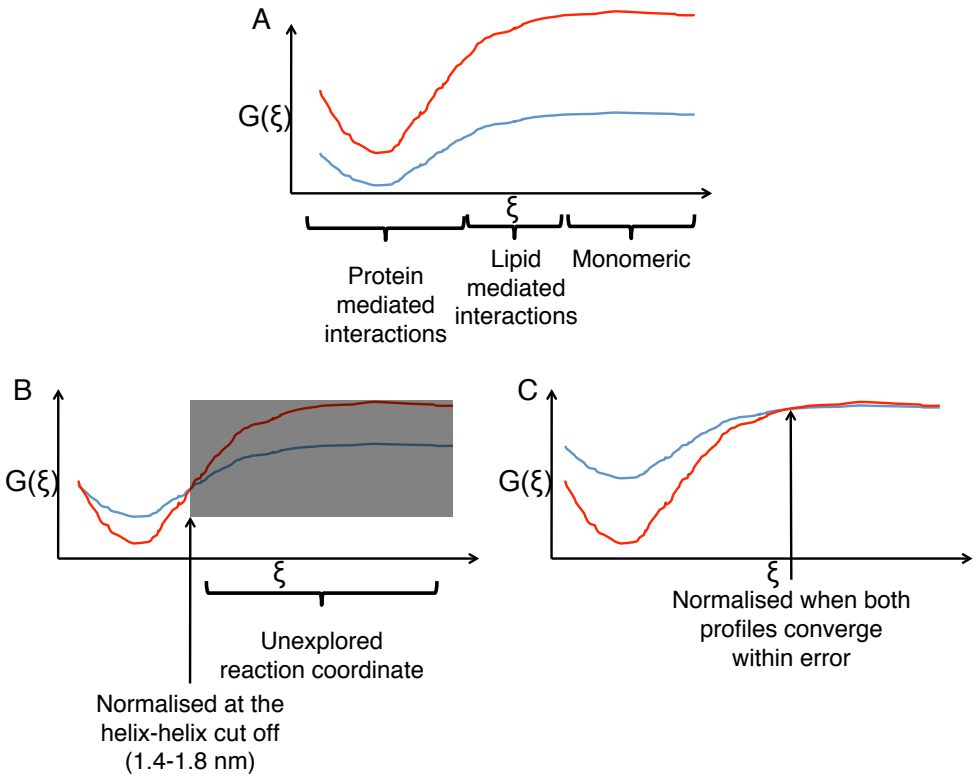

Figure 6.3: Typical PMF profiles of TM domain association. (A) Two PMF profiles before normalisation. (B) Normalisation of the two profiles where G(ξ) has not plateued. (C) Normalisation well past the the point at which G(ξ) has reached a plateau.

The technique of normalising PMF profiles has varied between computational studies. Dissipative particle dynamics (DPD) simulations of a CG model protein in DMPC by Yiannourakou *et al.,* (250) report an increase in free energy along an *inter*helical reaction coordinate within the short range interaction potential cut-off. All PMF profiles were normalised to a reference point of 0 kJ mol^{-1}. A further DPD study of the model WALP23 peptide in a DPPC bilayer by Al-Lehyani *et al.,* (251) normalised PMF profiles to 0 kJ mol^{-1} at the furthest value along the reaction coordinate

even though the PMF profile had not converged. Acknowledged in their own work, their study is limited by the system box size. Normalisation of comparative PMF profiles at a distance along an *inter*helical reaction coordinate as a response to a limited box size is also seen in the work of Polyansky *et al.* (96). The *inter*helical PMF profiles of eighteen united-atom encoded TM dimers were calculated across three different compositions of membrane bilayer. With an *inter*helical reaction coordinate of only 20 Å, profiles were not normalised to a converged reference point. The resulting ΔΔG between profiles may prove to be highly inaccurate given that the *inter*helical reaction coordinate distance was too short to account for any lipid mediated long range interactions.

This chapter acts as a bridge between the PMF profiles of the united atom Neu simulations in chapter 4 and the PMF profiles of the CG low complexity scaffolds in chapter 7. In the united-atom case, PMF calculations of Neu and Neu* continued to demonstrate an increase in ΔG (Figures 4.4), up to 6 nm along the *inter*helical reaction coordinate (Figures 4.9, 4.11, 4.10 and 4.12). The annular and nearby lipids were shown to order due to the positive mismatch between peptide and bilayer. The PMF profiles only converged once the set of ordered lipids surrounding each peptide were far enough apart so as to introduce bulk lipids in between. On the other hand, the PMF profiles of low complexity scaffolds (for example, Figures 7.1 and 7.11) clearly converge at a much shorter distance along the *inter*helical reaction coordinate. In light of the difference between the PMF profiles of the two force fields used in this study, and the reports in literature, it was of interest to understand whether these differences were a consequence of: the model (CG vs united atom); the TM domain sequence; or the TM domain length. The differences in free energy values and distance of free energy convergence between CG and united atom models were tested by comparing CG Neu_{WT} PMF profiles from section 4.3.2. To reliably compare PMF profiles from Neu_{WT} CG to Neu_{WT} united atom models, the structure of the idealised helical CG representation was optimised by using the angle distributions from the united atom simulations of a single Neu_{WT} peptide. The monomeric Neu_{WT} from the free energy calculations of chapter 4 was used. We did not generate a united atom model of a monomeric AZ2 peptide in order to optimise the CG AZ2 structure. To test the effects due to hydrophobic length, CG AZ2 dimers varying in hydrophobic varying between 22 to 32 hydrophobic residues were simulated in a CG POPC bilayer. This was tested further by comparing the tilt angle and bilayer thickness of the CG Neu_{WT} to the set of CG AZ2 dimers to elicit potential changes to bilayer thickness mediated from the TM sequence.

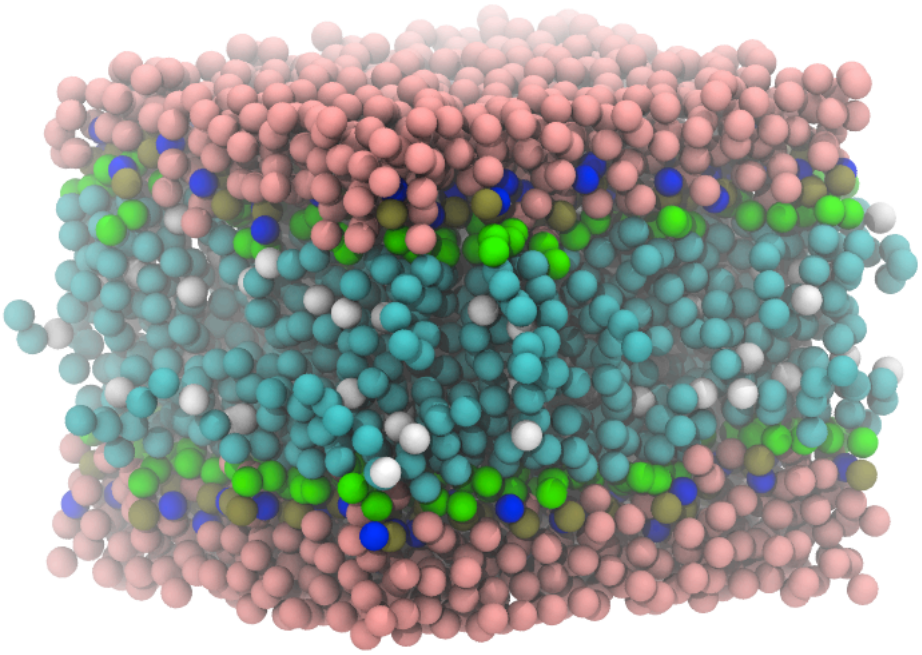

Figure 6.4: Final frame of a 500 ns simulation of a coarse-grain POPC lipid bilayer fully solvated in coarse-grain water molecules (pink beads). Most of the water molecules were removed for clarity.

6.2 Simulation methodology

6.2.1 Force field parameters

The peptides, POPC lipids, water and counter-ions were modelled using the CG MARTINI force field. A DPPC lipid was adjusted to incorporate the oleoyl carbon double bond as present in a POPC lipid. The 51 atoms from a united atom POPC lipid were amalgamated into 13 CG beads. All peptide models were built using the MARTINI toolkit (123). The hydrophobic TM domain of the peptide was defined as α-helical. The last three residues at either terminus were defined as a coil to simulate the non-helical secondary structure of the linker regions either side of the TM domain of the GALLEX protein used elsewhere in this study. Counter-ions were used to neutralise the charge on the arginine side chains.

6.2.2 Simulation parameters

Simulations were performed using Gromacs 4.5.5 (190). All bonds were constrained using the LINCS algorithm (155) and a time step of 0.025 ps was employed. The temperature was maintained at 323 K using the Berendesen thermostat (149) during the initial 50 ns equilibration before switch-

ing to the Nosé-Hoover thermostat (150) for further equilibration and production runs. Previous attempts at equilibrating the system using the Parrinello-Rahman pressure-coupling resulted in uncontrolled oscillation of the unit cell. Therefore, the Berendsen barostat pressure-coupling (149) scheme was used to maintain the pressure at 1 bar for an initial 50 ns equilibration before switching to the Parrinello-Rahman pressure-coupling scheme (153) for production simulations. The pressure-coupling was applied in a semi-isotropic manner, whereby the *z*-dimension of the simulation cell (in the direction of the bilayer normal) was scaled independently from the *xy* dimensions (in the plane of the bilayer). Long-range electrostatic interactions were treated using the particle mesh Ewald approach (127) with a short-range cut off of 1.2 nm. Lennard-Jones interactions were cut off at 1.2 nm and terminated by shifting the cut off to 0 kJ mol^{-1}.

6.2.3 POPC lipid bilayer construction

A single CG POPC lipid was replicated two hundred times in the *xy*-dimension to form a monomer. The second leaflet was constructed by copying the first and then inverting the structure. The structure was solvated with 10764 MARTINI water molecules. The final structure was energy minimised using the steepest descent algorithm (see section 2.4.1). To ensure the bilayer had relaxed, the area-per-lipid, pressure and temperature were observed during a 200 ns simulation using the simulation parameters outlined in section 6.2.2.

6.2.4 Insertion of dimers into a POPC lipid bilayer

Final frames from the prepared bilayer as described in section 6.2.3 were overlaid with a copy of the AZ2 dimers as listed in Table 6.1. Dimers were orientated into a left-handed helix-helix crossing angle before their centre of mass was positioned at the same central position of the POPC bilayer. The CG homo-dimers were inserted into the equilibrated POPC bilayer using the InflateGro tool as detailed in section 4.2.5. Once inserted, the system was left to equilibrate for a further 200 ns using an *NPT* ensemble.

6.2.5 Umbrella sampling and free energy calculations

A reaction coordinate was defined between helices by applying a 0.001 kJ mol^{-1} pull force to one helix in the *x*-dimension for 200 ns whilst restraining the second helix with 1000 kJ mol^{-1}

restraints on the backbone particles. Individual umbrella windows were generated from the steered MD simulations every 0.5 Å along the reaction coordinate. An umbrella potential of 800 kJ mol^{-1} was applied to the *xy*-lateral distance between the centre-of-mass of each helix. Each umbrella simulation ran for 1 μs and data points were collected between 0.1 μs to 1 μs to generate a PMF profile using WHAM (198). The quality of the reaction coordinate coverage was analysed by sampling over increments of 0.1 μs of data points to check for PMF convergence. Bayesian bootstrapping was used to calculate statistical uncertainty over 200 bootstrap iterations and the average free energy was taken as the final result. The free energy profile was normalised to 0 kJ mol^{-1} where there was no longer a change in free energy.

6.2.6 Optimisation of the Neu_{WT} CG force field

A bond angle distribution is controlled by the force restraint. With a strong force restraint the bond angle distribution is more frequent about its desired angle. Bond angle distributions can be captured over the course of a simulation, and the standard deviation of the bond angle frequency can be used to measure the effects a particular force constraint has on a bond angle. The CG MARTINI force field for Neu_{WT} was optimised to approximate the bond angle distributions produced from 50 ns NPT simulations of a united atom representation of Neu_{WT} using in-house software (designed and developed by the author of this thesis) based on a Genetic Algorithm scheme (252). The angle distributions from the united atom force field were used as a set of target bond angle distributions. An initial *population* of 100 sets of pure CG MARTINI bond angles and force restraints inside the forcefield input file were randomly perturbed by ±2° and 100 kJ mol^{-1}, respectively. A complete set of angles and force restraints is referred to as an *individual*. For each *individual*, an NPT simulation was performed for 50 ns and a new set of angle distributions and standard deviations were calculated. A cost function was assigned over the bond angle distributions, according to the equation:

$$cost = \frac{1}{n_g} \sum_{i=1}^{n_g} (\frac{1}{e^q}) tan(q), \tag{6.1}$$

where n_g represents the combined number of bond angles and force restraints per individuals, and:

$$q = |g_{ai} - g_{ei}|, \tag{6.2}$$

where g_{ai} represents the actual bond angle from the generated bond distribution as a result of the simulations, and g_{ei} represents the united atom target bond angle. A score is assigned to each individual according to the inverse of the cost:

$$fitness = \frac{1}{score} \tag{6.3}$$

The *individual* with the most desirable set of bond angle values will have the highest $fitness$ score amongst the population. In order to refine the bond angle values further, the fittest 25 *individuals* were retained. The remaining population is arranged into pairs. Each pair of *individuals* is split in two at a random within both *individual* sets. Their content is then switched resulting in two new *individuals*. This is commonly referred to as *crossover*. After the depleted population had been restored, each bond angle and standard deviation was perturbed 25% from the desired values on a probability of 0.05 %. This technique is commonly referred to as a *mutation* operation, and is used to present variation in the otherwise limited search space. Given the new population of bond angle and force restraints, a further round of 50 ns NPT simulations are performed. The fitness is recalculated, weak *individual* are discarded and the fitter *individual* are retained to generate a new population. This continued until 10 successive populations continued to yield 25 *individual* whose total fitness only deviated by 2 %. The fittest *individual* from the final population is used to perform a 500 ns NPT simulation.

6.3 Results

6.3.1 Optimisation of the Neu_{WT} CG force field

The complete set of bond angles were obtained from a 500 ns non-optimised MARTINI model simulation, the united atom model simulation, and the optimised MARTINI model simulation, notably: 31 backbone-backbone-backbone angles (Figure 6.5 (A)); 9 backbone-side chain-side chain angles (Figure 6.5 (B)); 9 backbone-backbone-side chain angles (Figure 6.5 (C)); and 24 backbone-backbone-backbone-backbone dihedral angles (Figure 6.5 (D)). The recorded angle

distributions suggest that the initial CG representation of the Neu_{WT} peptide does not approximate the angle distributions obtained from the united atom reference simulation in almost every case. The angle distributions from the simulation using the optimised MARTINI force field yields an accurate representation of the angle distributions from the simulations using the united atom force field.

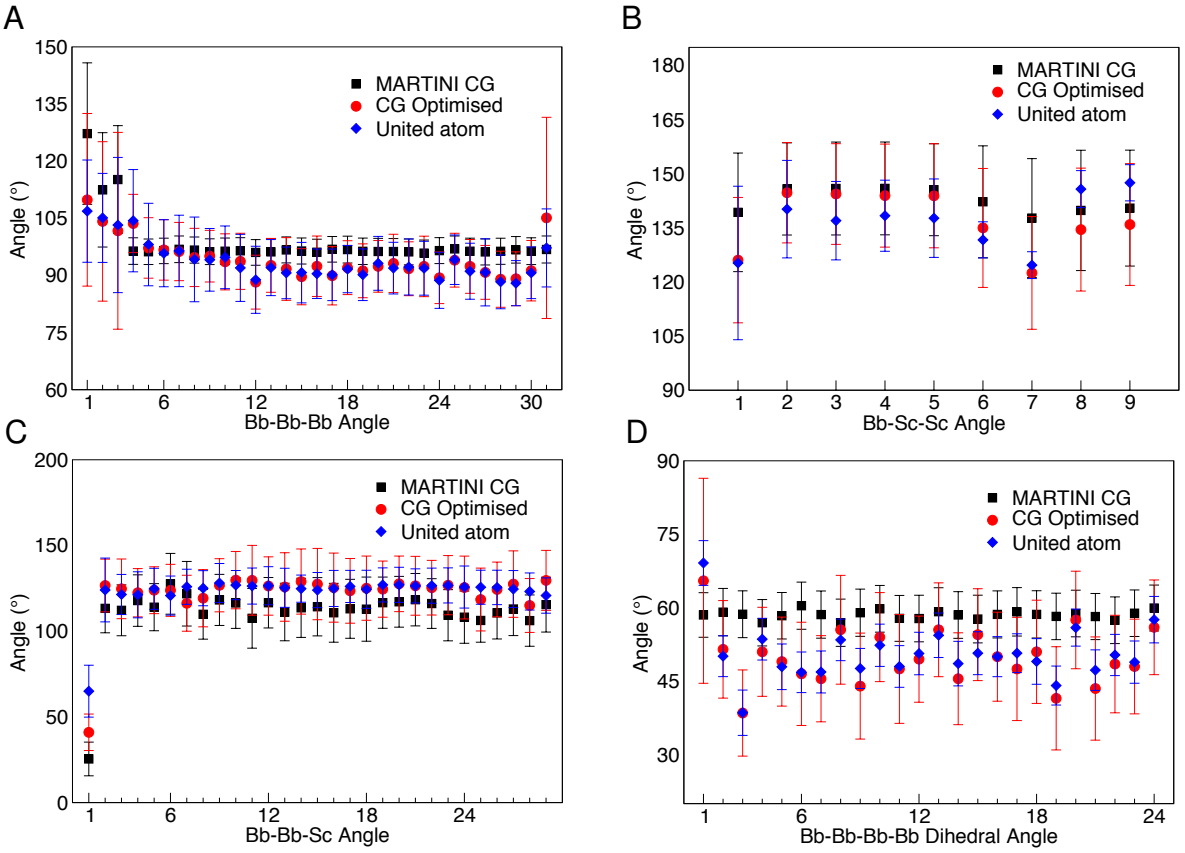

Figure 6.5: Comparison between the bond angle distributions of the force fields: MARTINIC CG; united atom; and optimised CG. (A) Backbone-backbone-backbone angle distributions, (B) backbone-side chain-side chain angle distributions, (C) backbone-backbone-side chain angle distributions, and (D) backbone-backbone-backbone-backbone proper dihedral angle distributions. The standard deviation represents the flexibility in the distribution of each angle.

6.3.2 Free energy of association between alanine zippers of varying hydrophobic lengths

PMF profiles of $AZ2_{22}$, $AZ2_{32}$ and Neu_{WT} were calculated using umbrella sampling and WHAM. Helices were separated by 4 nm along an *xy*-plane *inter*helical reaction coordinate. Histograms constructed from 0.1 μs to 1 μs of umbrella sampling data points, as seen in Figures 6.6 (A), 6.7 (A), and 6.8 (A), showed adequate overlap between histograms.

Final 0.1 μs to 1 μs PMF profiles were checked for sufficient sampling by constructing a series of

incremental profiles starting from 0.1 μs. As seen in Figures 6.6 (B), 6.7 (B), and 6.8 (B), final 0.1μs to 1 μs profiles either converged beside incremental PMF profiles, or fell between the profiles of two previous PMF profiles. Both would suggest that the umbrella window trajectories had sampled enough data points across the reaction coordinate.

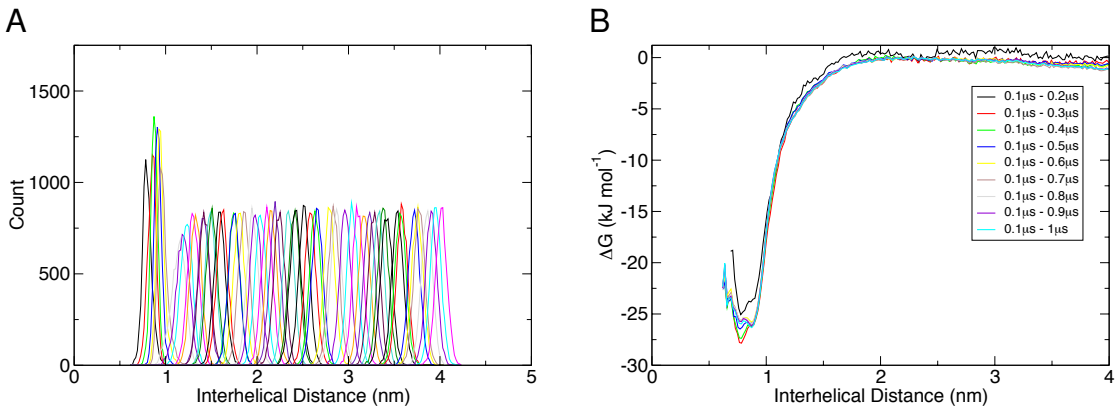

Figure 6.6: A) Umbrella sampling histogram coverage along an *inter*helical distance reaction coordinate and (A) block averaging PMF profiles, of $AZ2_{22}$.

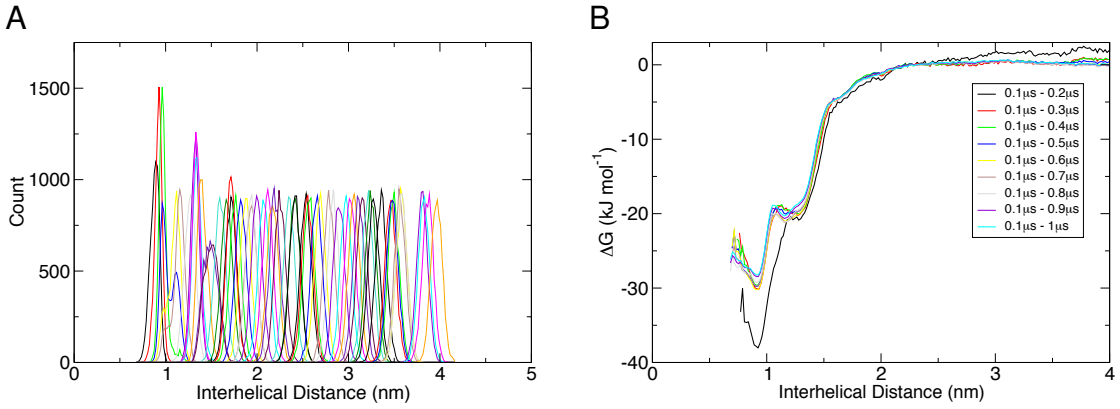

Figure 6.7: (A) Umbrella sampling histogram coverage along an *inter*helical distance reaction coordinate and (A) block averaging PMF profiles, of $AZ2_{32}$..

As shown in Figure 6.9, The free energy of self association is approximately -26.5 kJ mol^{-1} at 0.85 nm and -28 kJ mol^{-1} at 0.9 nm for $AZ2_{22}$ and $AZ2_{32}$, respectively. Both $AZ2_{22}$ and $AZ2_{32}$ PMF profiles converge between 2 nm and 2.5 nm along the *inter*helical reaction coordinate. The free energy of self association of Neu_{WT} is approximately -41.5 kJ mol^{-1} at 0.72 nm along the reaction coordinate. There is a local minima 1.15 nm along the reaction coordinate, with a free energy -37 kJ mol^{-1}. The Neu_{WT} PMF profile converges 2.75 nm along the *inter*helical reaction coordinate. This is, at most, 0.75 nm further along the reaction coordinate than either AZ2 PMF profiles. All three PMF profiles remain within error for the remainder of the reaction coordinate after convergence.

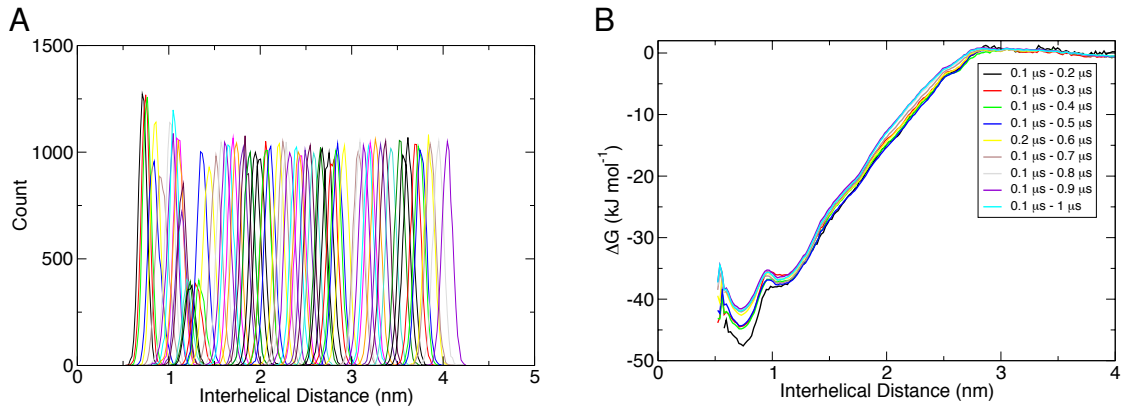

Figure 6.8: A) Umbrella sampling histogram coverage along an *inter*helical distance reaction coordinate and (A) block averaging PMF profiles, of Neu_{WT}.

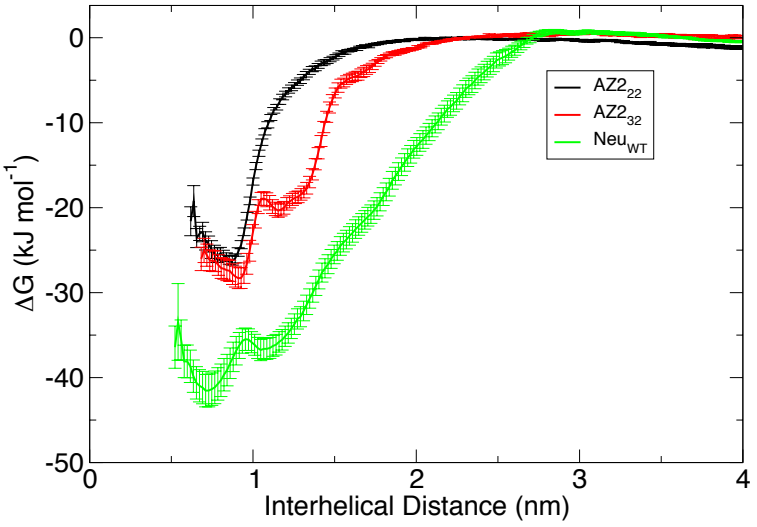

Figure 6.9: The potential mean force profile along an *inter*helical distance reaction coordinate for the TM domains $AZ2_{22}$, $AZ2_{32}$. and Neu_{WT}. PMF profiles were normalised to 0 kJ mol^{-1} at a reaction coordinate value of 2.7 nm.

6.3.3 Bilayer thickness of a CG POPC lipid bilayer

The average bilayer thickness in the absence of TM proteins was calculated from a 500 ns NPT ensemble production run using the GridMAT-MD tool as previously parameterised in section 4.3.1.1. The average bilayer thickness between opposite leaflet CG phosphate groups can be seen in Figure 6.10. The standard deviation represents the fluctuations in the bilayer. The overall average bilayer thickness was approximately 4.21 ± 0.28 nm, with very little in the way of observable fluctuations in thickness or excessive curvature. The area per lipid of bulk POPC was 67.96 ± 0.71 $Å^2$, which compares well with experimental approximations (253).

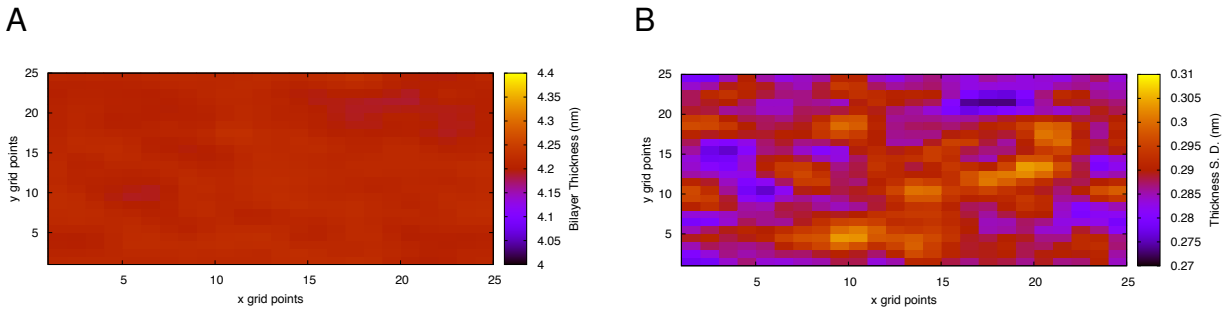

Figure 6.10: (A) Average CG POPC bilayer thickness over a 500 ns NPT production run. (B) Standard deviation of the bilayer thickness reported at each grid-square.

6.3.4 Bilayer thickness and area per lipid of mediated through hydrophobic TM domain length

Each CG AZ2 dimer and Neu_{WT} dimer was inserted into a CG POPC lipid bilayer and left to equilibrate as described in section 6.2.4. The simulation time was sufficient to capture bilayer physiological effects and changes to dimer orientation. During the course of the simulation, each set of helical TM peptides remained dimeric and were not expelled from the bilayer. The bilayer thickness was calculated as per section 4.3.1.1 and are presented along with the final frame from each trajectory (Figure 6.11). An overall average of the bilayer thickness is presented in Table 6.1.

The presence of an $AZ2_{22}$ dimer results in periodic curvature to the bilayer either side of the dimer (Figure 6.11 (A)). However, the standard deviation of bilayer thickness (Figure 6.13) would suggest that the observed curvature in addition to the average bilayer thickness of all CG simulations were statistically insignificant. Further to this point, the bilayer thickness of $AZ2_{24}$, $AZ2_{26}$ and $AZ2_{32}$ were approximately 4.3 to 4.4 nm, and $AZ2_{28}$ and $AZ2_{30}$ were approximately 4.1 to 4.35 nm, and the standard deviation would suggest that the bilayer thickness values were not statistically significant.

Projection of the bilayer average onto the bilayer length (Figure 6.13), clearly suggests that there is no statistical difference between the bilayer thickness in the presence of any AZ2 dimers. Neither is there a significant difference between the bilayer thickness of each system and the bulk POPC bilayer thickness. The introduction of a Neu_{WT} dimer into a POPC bilayer resulted a drop to the average bilayer thickness compared with the bilayer thickness of bulk POPC. However, not only was the bilayer thickness statistically equivalent across the whole bilayer (Figure 6.12 (A) and (B)),

but it was comparable to the bilayer thickness results of all AZ2 system. This suggests that the change in difference in TM sequence failed to induce a change to the bilayer structure, and that the combined hydrophobic length of the Neu_{WT} dimer along with its sequence did not cause a difference to the thickness of lipids far from the dimer.

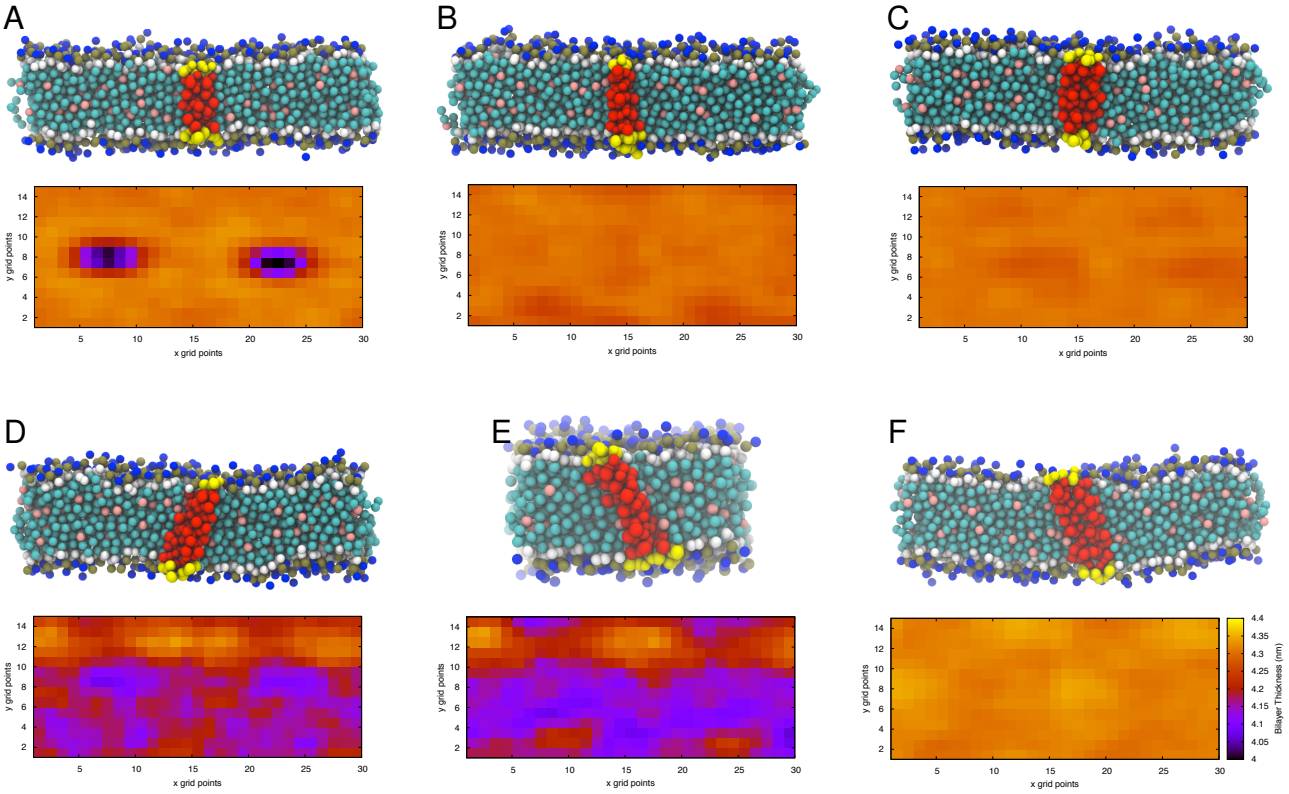

Figure 6.11: Final frame representation of a 500 ns POPC with an AZ2 dimer varying in TM domain length along with an average bilayer thickness surface plot. (A) $AZ2_{22}$, (B) $AZ2_{24}$, (C) $AZ2_{26}$, (D) $AZ2_{28}$, (E) $AZ2_{30}$, and (F) $AZ2_{32}$. Figure (E) was captured along the adjacent orientations to the remaining figures so as to capture the effects to the dimer tilt angle.

6.3.5 Tilt angle mediated through hydrophobic TM domain length

The average hydrophobic span of the AZ2 dimer increases as the heptad repeat sequences is extended with additional alanine/leucine residues. The fluctuating hydrophobic lengths of each dimer (Figure 6.14 and Figure 6.15), calculated as the average distance between the first and last hydrophobic residue across both helices, are due to very small adjustments to their helix-helix packing alignment. In addition, given that the MARTINI forcefield prevents the peptides from a protein's secondary structure, a large change to the dimer hydrophobic length would not have been expected.

As the hydrophobic length of the TM dimer is increased, it was expected that eventually the

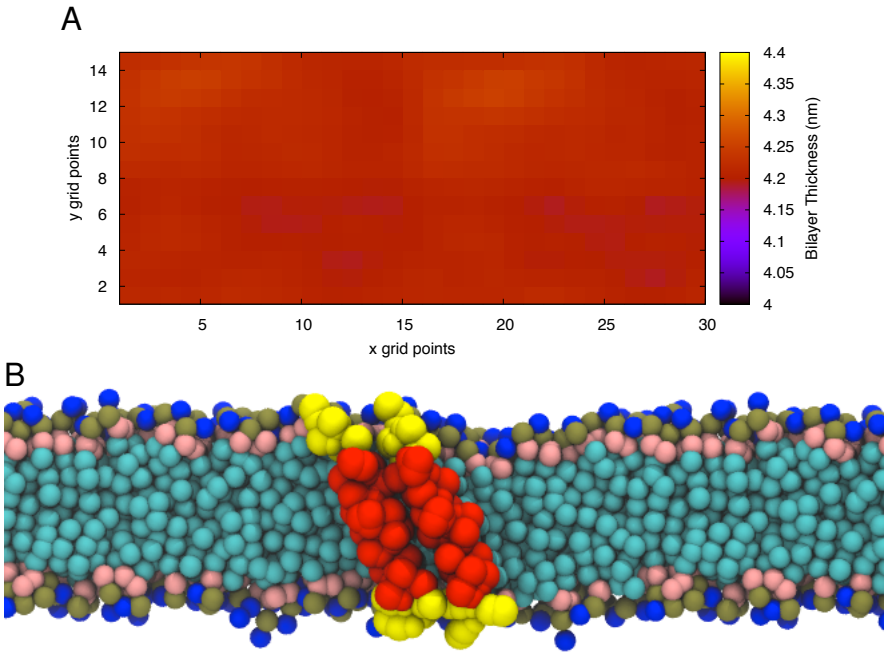

Figure 6.12: (A) Averaged CG POPC bilayer thickness in the presence of a CG Neu_{WT} dimer. (B) Snapshot of the final frame from a 500 ns simulation of CG Neu_{WT} dimer in POPC. The red residues signify the hydrophobic peptide span measured from F_{658} to I_{680}. Hydrophilic at both ends are represented in yellow. Water has been removed for clarity.

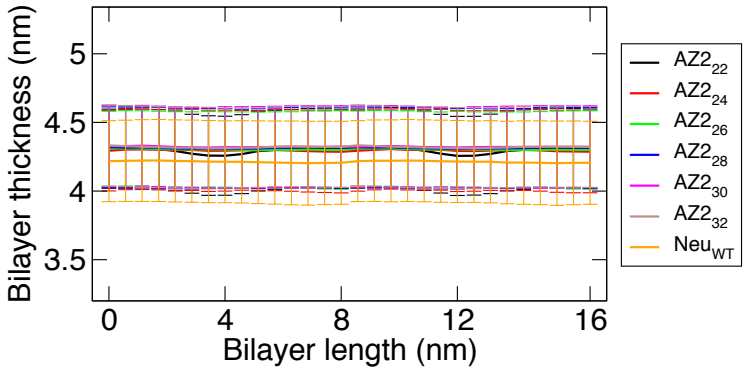

Figure 6.13: Cross sectional averaged bilayer thickness calculated across the length of the bilayer. Error bars represent standard deviation.

relationship between dimer and bilayer would experience a positive hydrophobic mismatch resulting in an adjustment to the dimer tilt angle (Figure 6.2 (B)). The angle between the principle axis of a dimer and the bilayer normal was calculated from the recorded trajectories. As surmised in Table 6.1, and given as a frequency distribution (Figure 6.16), the first four AZ2 dimers: $AZ2_{22}$; $AZ2_{24}$; $AZ2_{26}$; and $AZ2_{28}$, return very similar dimer tilt angles within error, between 9.15° to 11.51°. The hydrophobic span of these dimers is between 2.42 nm to 3.46 nm (see Figure 6.14), which falls shorts of a bulk POPC average bilayer thickness of 4.21 nm ± 0.28 nm.

There is a slight increase in tilt angle of $AZ2_{30}$, reporting an average of 14.79° ± 7.23°. The

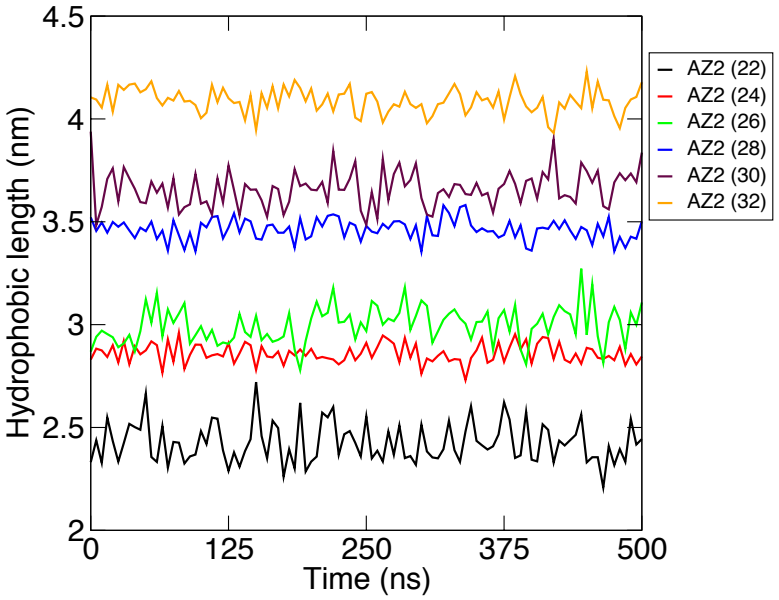

Figure 6.14: Averaged AZ2 dimer hydrophobic lengths as a function of time. The average hydrophobic length increases as additional hydrophobic residues are included.

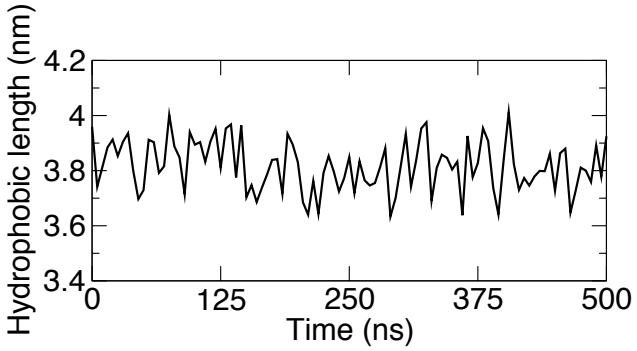

Figure 6.15: Neu_{WT} dimer hydrophobic length across a 500 ns production simulation..

hydrophobic span of the $AZ2_{32}$ dimer, at 4.1 ± 0.1 nm, is within reach of the hydrophilic lipid head group of bulk POPC. Given this increase in hydrophobic length, it is not surprising to find that $AZ2_{32}$ returned an average tilt angle of approximately 26± 7°. It is surprising, however, that the bilayer did not increase in thickness as it had been observed from the united atom simulations of Neu in section 4.3.3.

The hydrophobic span of the Neu_{WT} dimer was along the average length of 24 hydrophobic residues, sharing the same number of hydrophobic residues as $AZ2_{30}$. However, due to the difference in helical turn due to the inclusion of a large number of valine side chains, Neu_{WT} is approximately 0.44 nm longer in the dimeric hydrophobic length. With the increase in hydrophobic span, it is of

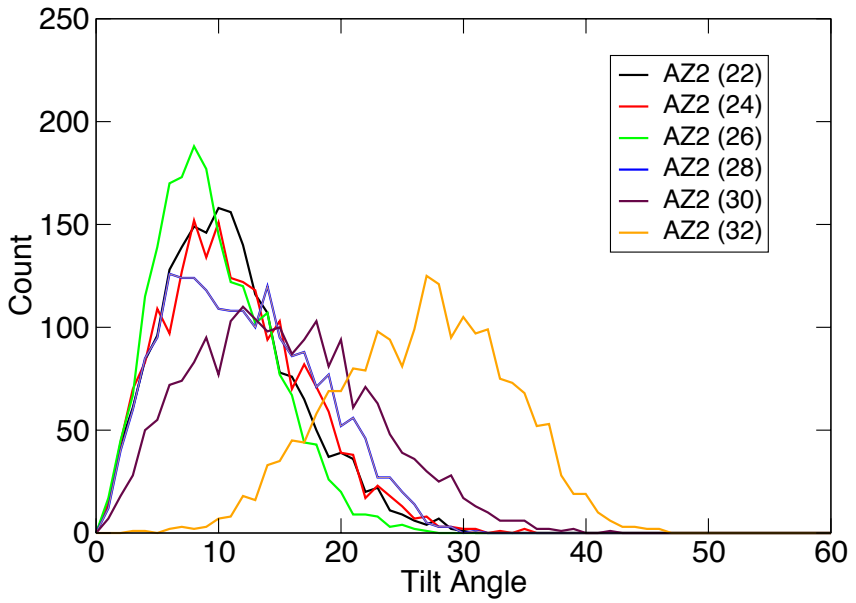

Figure 6.16: $AZ2_{22}$ through to $AZ2_{32}$ dimer tilt angle across 500 ns production simulations.

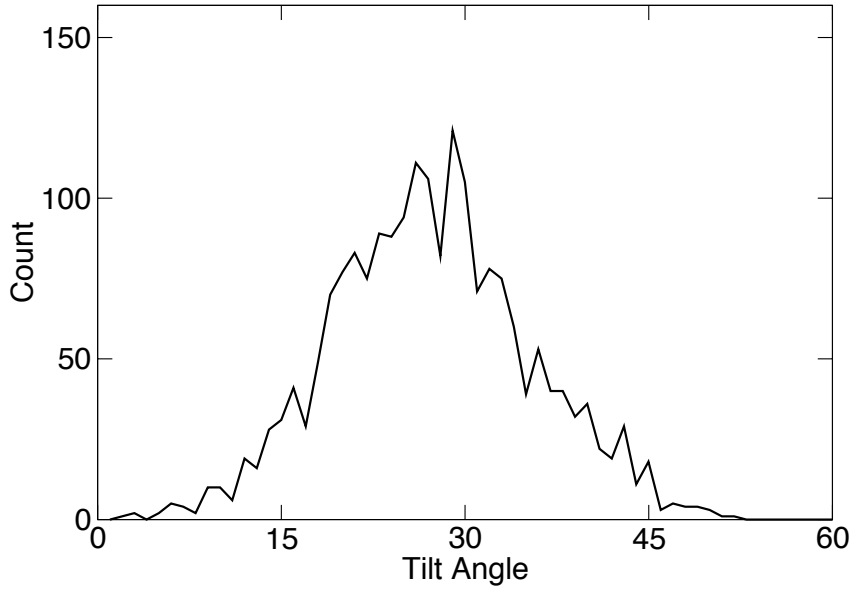

Figure 6.17: Neu_{WT} dimer tilt angle across a 500 ns production simulation.

no surprise that Neu_{WT} adopts a greater dimer tilt angle than $AZ2_{30}$. Neu_{WT}, although shorter in hydrophobic residues, is more comparable in hydrophobic length to $AZ2_{32}$, seemingly adopting a similar dimer tilt angle of approximately 26° within error.

6.4 Discussion

The aim of this chapter was to determine whether a CG TM dimer embedded in a CG POPC bilayer could reproduce the hydrophobic mismatch between bilayer and peptide as seen in the united atom PMF profiles of Neu, in chapter 4. This study used a set of CG low complexity scaffolds (the heptad sequence AZ2 from chapter 7), varying in hydrophobic length, and a CG Neu_{WT} dimer. To begin with, the bond angle distributions from a 500 ns MD simulation of a MARTINI force field parameterisation of Neu_{WT} using the NPT ensemble, did not compare well to the bond angle distributions of a single Neu_{WT} united atom simulation. Therefore, the force field for the CG Neu_{WT} peptide was optimised by iteratively adjusting the bond angles and force constraints using a Genetic Algorithm.

Using the optimised Neu_{WT} structure and the AZ2 dimers, bilayer thickness, dimer tilt angle and dimer hydrophobic length were calculated. To compare the affects on PMF calculations, umbrella sampling was performed on the dimer with the shortest and longest hydrophobic length, $AZ2_{22}$ and $AZ2_{32}$, respectively, and on the Neu_{WT} dimer. Neither topological or cross-sectional averaging of the bilayer thickness over the 500 ns MD simulations of AZ2 and Neu_{WT} dimers presented a discernible trend in changes to the bilayer thickness as a function of dimer hydrophobic length. The reported average thickness of $AZ2_{24}$, $AZ2_{26}$, and $AZ2_{32}$ returned a relatively even average bilayer thickness of approximately 4.3 nm to 4.4 nm. Taking into account the effects of a periodic simulation cell, $AZ2_{28}$ and $AZ2_{30}$ form a depression across the length of the unit cell, and at least half as wide. There is a clear difference between the depression and the elevated bilayer regions of approximately 4.15 nm to 4.4 nm, with the dimer positioned within the depression. $AZ2_{22}$ on the other hand, clearly demonstrate a depression in two regions of the bilayer, either side of the dimer. This would suggest, that the length of the $AZ2_{22}$ hydrophobic TM dimer region were within negative hydrophobic mismatch of the hydrophobic region of the lipid bilayer, a consequence of the three hydrophilic residues either end of each peptide coming into contact with the lipid tails. Interestingly, the average cross-sectional bilayer thickness of the Neu_{WT} system was less than $AZ2_{22}$, even though the Neu_{WT} hydrophobic span of residues was longer. Unfortunately, unlike the united atom Neu simulations in chapter 4, none of the bilayer thickness calculations were statistically significant, in either topological or cross-sectional form, implying that the selected

range of hydrophobic dimers were no noticeable affect on the bilayer thickness. Earlier work on lipid bilayer perturbation from the presence of TM domain peptides and nano-structures were investigated to determine whether the work in this study was an isolated case. According to Nielsen *et al.,* (254), using a pre-MARTINI CG force field representation to parameterise a hydrophobic nanotube in a DMPC bilayer they investigated the alleviation to hydrophobic mismatch as a function of nanotube width. Interestingly, they were able to capture changes to the bilayer thickness as a function of distance from the centre of mass of the nano tube as the width of the nanotube was increased. Venturoli *et al* (249), were also able to capture changes to bilayer thickness around a hydrophobic peptide in a mesoscopic scaled system.

In addition to a change in the bilayer thickness, the dimer can adjust by increasing or decreasing the dimeric tilt angle. Nielsen *et al.,* (254) concluded that a narrow hydrophobic TM structure would alleviate hydrophobic mismatch mainly through altering their tilt angle, with a slight swelling to the neighbouring bilayer region as the TM structure for very long TM structures. As the width of a TM structure increased, there was less of a change to the tilt angle and more of a thickening of the local lipid regions. Venturoli *et al* (249) hypothesised that wide TM nano-structures relied on deformation to the bilayer as tilting was energetically unfavourable compared to the energetic expense to bilayer changes. Figure 6.18 compares the dimer tilt angle with the dimer hydrophobic length of the AZ2 dimers from this study. It is clear that the hydrophobic length of $AZ2_{22}$ to $AZ2_{28}$ does not significantly change the tilt angle. However, the average tilt angle of $AZ2_{30}$ shows an average tilt angle increase, and finally $AZ2_{32}$ shows a statistical significant increase in tilt angle.

The tilt angle of CG Neu_{WT} was $27° \pm 8°$, very similar to the tilt angle of $AZ2_{32}$. This was expected given the similarity between their hydrophobic lengths. It was clear from careful observation of the CG Neu_{WT} trajectory that the peptides formed a dimer with the $I_{659}xxxV_{663}$ interaction motif at the helix-helix interface. A direct comparison between the CG and united atom tilt angles is difficult to make. Given that the united atom simulations were restrained using an umbrella potential along a reaction coordinate and the simulation time was 40 ns, the CG simulation may have found an alternative free energy minimum in the simulation time given. In addition, the united atom Neu_{IV} yields a left handed helix-helix crossing angle at the global free energy minimum, where as the CG Neu helix-helix crossing angle is close to perpendicular. The orientation of a helix-helix crossing angle can significantly affect the free energy of TM dimer self-association as the range of enthalpic

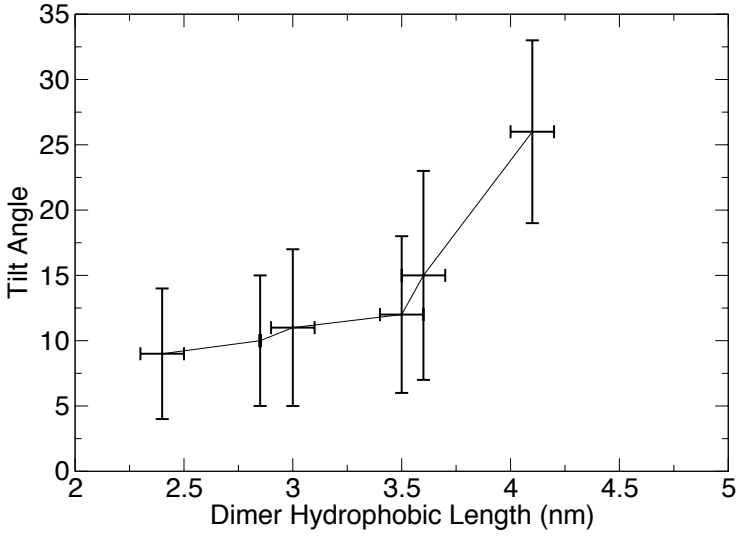

Figure 6.18: CG AZ2 tilt angle as a function of dimer hydrophobic length.

contributions from side chains would vary (8).

The average global free energy minimum of CG Neu_{WT} was -41 kJ mol^{-1}, which is in close agreement with the free energy of self-association of CG ErbB-2 by Prasanna *et al* (216). In addition, their PMF profile converged at a reaction coordinate value of 2.2 nm, which is in very good agreement with the CG Neu_{WT} PMF profiles presented in this study. The PMF profiles of CG Neu_{WT} compared with the two AZ2 dimers would suggest that the location of the free energy plateau is to a degree, sequence dependent given the similarity to the hydrophobic length and tilt angle of $AZ2_{32}$. When compared with the united atom PMF profile, as seen in Figure 6.19, the plateau of the CG Neu_{WT} starts considerably earlier, by as much as 4.2 nm along the reaction coordinate, compared with the united atom simulation. There is an approximate difference in free energy of self-association by 67 kJ mol^{-1} between CG Neu and the united atom Neu_{IV} PMF profiles. We believe that this difference is due to the limited changes to the bilayer which mediate the long range effects seen in the chapter 4.

6.5 Conclusion

According to the lack of significant change to the bilayer thickness of the simulations presented in this chapter, the MARTINI CG model does not accurately simulate the changes to the bilayer thickness in the presence of a Neu homodimer as seen in the earlier united atom representation

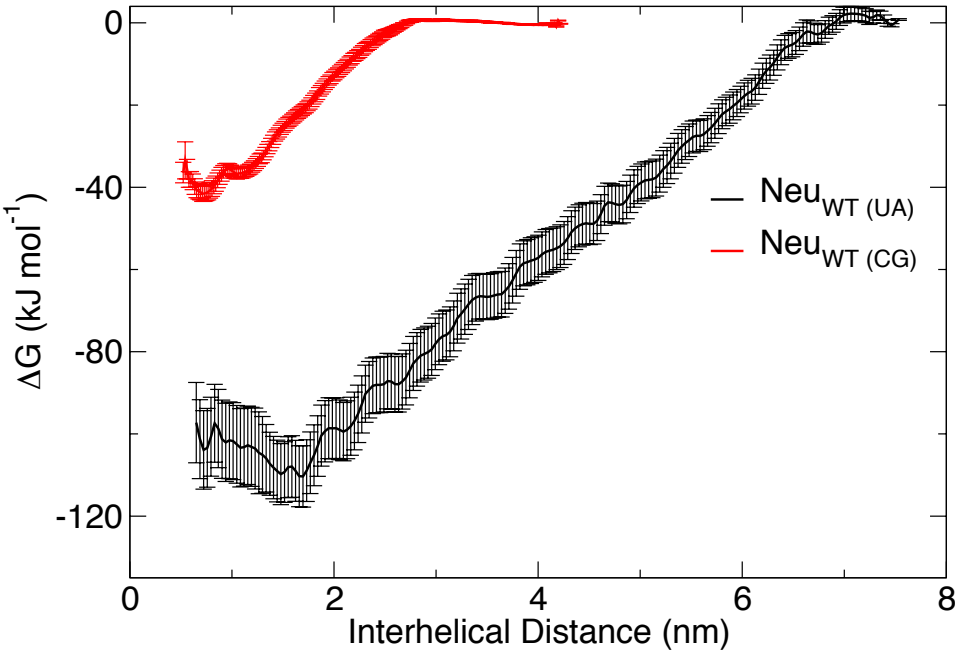

Figure 6.19: Neu_{WT} coarse-grain (CG) and united atom (UA) PMF profiles. Complete profiles were normalised to 0 kJ mol^{-1} at the location in where they began to converge.

(chapter 4). This may be due to a significant reduction in the number of degrees of freedom in the lipid CG model. However, the changes to the CG AZ2 dimer tilt angles as a function of hydrophobic length does suggest that the CG simulations are modelling a degree of hydrophobic mismatch even though changes to the bilayer thickness were not reported. It is for this reason why we believe that the PMF profiles of CG TM dimers in this study are converging by as much as 4.2 nm earlier than the united atom PMF calculations. However, as presented in the discussion, there is sufficient evidence in the literature to suggest a CG model can capture deformation to the bilayer providing the length and/or width of the TM structure is sufficient enough. A final step would be to perform similar simulations over a variety of CG force fields, which were not included in this study.

Peptide	N-terminus	Hydrophobic stretch	C-terminus	Length of hydrophobic stretch (nm)*	Dimer tilt angle (°)	Area per lipid (Å^2)	Bilayer thickness (nm)
Bulk POPC						68 ± 0.8	4.2 ± 0.3
$AZ2_{22}$	Ac-GSS	$(ALLAALL)_2$-AL	$TSR\text{-}NH_2$	2.4 ± 0.1	9 ±5	65 ± 1	4.3 ± 0.3
$AZ2_{24}$	Ac-GSS	$(ALLAALL)_2$-ALLA	$TSR\text{-}NH_2$	2.85 ± 0.04	10 ± 5	64.19 ± 0.93	4.3 ± 0.3
$AZ2_{26}$	Ac-GSS	$(ALLAALL)_2$-ALLAAL	$TSR\text{-}NH_2$	3.0 ± 0.1	11 ± 6	64 ± 1	4.3 ± 0.3
$AZ2_{28}$	Ac-GSS	$(ALLAALL)_3$-A	$TSR\text{-}NH_2$	3.5 ± 0.1	12 ± 6	64.42 ± 0.94	4.3 ± 0.3
$AZ2_{30}$	Ac-GSS	$(ALLAALL)_3$-ALL	$TSR\text{-}NH_2$	3.6 ± 0.1	15 ± 8	64.35 ± 0.94	4.3 ± 0.3
$AZ2_{32}$	Ac-GSS	$(ALLAALL)_3$-ALLAA	$TSR\text{-}NH_2$	4.1 ± 0.1	26 ± 7	64.26 ± 1.64	4.3 ± 0.3
Neu_{WT}	Ac-RASP	VTFIATVVGVLLFLILVVVVGILI	$KRRR\text{-}NH_2$	3.8 ± 0.9	27 ± 8	68 ± 1	4.3 ± 0.1

Table 6.1: Bulk POPC, AZ2 and Neu_{WT} sequences, dimer hydrophobic length, dimer tilt angle, area per lipid and bilayer thickness. Standard deviation was used to calculate error.

Chapter 7

Molecular Dynamics simulations of low complexity TM domains

7.1 Introduction

The propensity for self-association between a set of low complexity scaffold TM domains before and after substitution with a helix-helix interaction motif was presented in chapter 5. In addition, a simplified monomer-dimer equilibria model was used to elicit the Gibbs free energy of TM domain oligomerisation in a biological context. The results so far present vital information on how these low complexity scaffold sequences behave *in vivo*. Unfortunately, beyond identifying which amino acid substitutions play a key role towards oligomerisation, it is not possible to analyse the relationship between structure and function from this data alone. Therefore, this chapter presents a computational approach to complement the aforementioned *in vivo* GALLEX results with free energy calculations, and *inter*helical side chain interaction analysis. There are three computational approaches taken in this study, a CG MD simulation, an atomistic conformational search using short MD steps, and a united-atom unrestrained MD simulation. Each approach is described below.

Firstly, low complexity scaffold TM domains described in section 5.1.1 were modelled using the CG Martini force field (123). Two helices of seventeen residues, capped with three residues at either terminal to model the hydrophilic GALLEX chimera linker regions, were embedded in a fully solvated CG POPC lipid bilayer. A POPC bilayer was selected, as its structural properties

have been parameterised well (194), with a bilayer thickness (37.9 Å) similar to the thickness of the E. coli inner membrane (195), enabling a comparison to out earlier *in vivo* studies. Umbrella sampling was performed along an *inter*helical reaction coordinate along the *xy*-dimension parallel to the bilayer and a PMF profile was constructed using the WHAM (198). The monomer-dimer equilibrium model of the GALLEX free energy calculations in a biological environment, see section 5.3.3, is a simplified model of what in truth is a multitude of equilibria as a result of a host of non-specific biological interactions between the TM protein of interest and other biological molecules. Therefore, a direct comparison between computational and experimentally obtained free energy calculations cannot be made. Yet, a comparison across the rank order of free energy calculations is still possible with the assumption that growth conditions of every cell sample were identical.

Secondly, to accompany the CG simulations, a conformational search of atomistic dimers in a low dielectric implicit environment was modelled using the CHI algorithm. Structures within a set RMSD were grouped into clusters. Although anisotropic characteristics of an explicit bilayer are absent, interaction energies of *inter*helical side chain interactions and side chain-implicit environment interactions can yield probably structures. In addition, unlike the CG model, an averaged structure from an identified cluster can model potential hydrogen bonds.

The trajectory pertaining to the global free energy minimum of self-association per low complexity scaffold was identified by calculating the average *xy*-planar distance between the centre of mass of each helix with the location of the global free energy minimum on the reaction coordinate. The protein atoms from the identified trajectory were subjected to single-linkage cluster analysis in order to establish the most prevalent oligomeric structure and helix-helix crossing orientation. Clusters from the CG low energy model and CHI conformational searches were used to construct a thermodynamic and structural pathway (see Figure 7.43): an extension to the pathway presented in the discussion of chapter 5.

Finally, the GALLEX analysis presented in chapter 5 suggests that a single serine residue in the context of a low complexity AZ2 sequence stabilises oligomerisation. On the other hand, Gray and Matthews (234) along with Ballesteros (235) suggested that a single serine residue would form an *intra*helical hydrogen bond, contributing less to oligomerisation than one would expect from a polar residue towards the centre of the bilayer core. Therefore to probe this result further, initial

structures of AZ2, AZ2L6S and AZ2L10S TM domains plus linker regions were built using the CHI algorithm and not subjected to a conformation search. Each atomistic dimeric structure was then converted into a united atom representation and inserted into a fully hydrated POPC bilayer. United atom simulations were performed in an NPT ensemble and the orientation of the serine side-chains were monitored along with the helix-helix crossing angle.

The single point glycine to isoleucine substitutions of L17(GG4), AZ2$(GG4)_L$, and AZ2$(GG4)_A$ were not modelled. Their sole purpose had been to confirm the *in vivo* contribution GG4 made to oligomerisation.

7.2 Simulation methods

The CG, CHI conformation search and united atom simulation parameters are largely identical to those presented in chapters 4 and 6, unless otherwise stated.

7.2.1 Coarse grained simulation configuration

CG free energy simulations were performed on low complexity scaffold pairs: L17; L17(GG4); L17L9Q; AZ2, AZ2$(GG4)_L$; and AZ2$(GG4)_A$.

7.2.1.1 Force field parameters

CG force field parameters were set to those previously defined in section 6.2.1. TM peptide sequences were seventeen residues in length (Table 7.1).

7.2.1.2 Simulation parameters

The temperature was set to 298 K to match the laboratory conditions of the assays presented in chapter 5. All other parameters are identical to those previously defined in section 6.2.2.

7.2.1.3 POPC bilayer construction

The construction of the POPC bilayer was the same as previously defined in section 6.2.3.

7.2.1.4 Insertion of dimers into a POPC bilayer

Dimers were inserted into pre-equilibrated POPC bilayers as previously defined in section 6.2.4

7.2.1.5 Umbrella sampling and free energy calculations

Umbrella sampling along an *inter*helical centre-of-mass reaction coordinate was identical as previously defined in section 6.2.5. CG PMF profiles of TM domain association converged between 2 nm to 2.3 nm along the reaction coordinate, therefore, umbrella windows were only defined up to 2.5 nm of the reaction coordinate. All PMF profiles were normalised to 0 kJ mol^{-1} at 2.3 nm. The harmonic potential applied to each umbrella windows was set to 1000 kJ mol^{-1}, yet in some cases it was reduced to 800 kJ mol^{-1} for additional umbrella windows. This ensured sufficient coverage of the reaction coordinate and overlap between neighbouring distributions.

7.2.1.6 Single-linkage cluster analysis

Single-linkage cluster analysis was performed over the trajectory that best matched the CG global free energy minimum. The first structure from each cluster was used as a representative structure. Each dimeric structure from a single time frame begins as its own cluster. Clusters are then iteratively joined to form larger clusters providing the RMSD between clusters is below 0.5 Å. This technique yields at most the same number of clusters as there are time frames (4000 per umbrella window simulation). Initial analysis yielded single-member clusters if the RMSD cut-off was below 5 Å. Clusters were identified for further analysis providing the cluster had a population of ten or more structures. This minimum cluster population was selected due to the propensity for single-linkage cluster analysis in this study to form a small number of very dominant clusters. Identifying clusters by using a smaller population cut off enables a better representation of the whole trajectory. The population in each identified cluster was reported as a percentage of the whole trajectory.

7.2.2 CHI conformational search

Unless otherwise stated, the centre-of-mass *inter*helical distance was set to 1.0 nm. The helices were rotated from 0° to 360° in increments of 30°. The cut off for the root mean squared difference between candidate structures was 0.1 nm. Up to eight structures within this cut off constitute a complete cluster. The atomistic positions of each member of the cluster are then averaged to yield a potential starting configuration for further MD simulations. All remaining parameters are identical to those defined in section 4.2.3. The initial dimer crossing angle, although not restrained by a

potential energy penalty, is very unlikely to adopt a statistically significant difference between clusters given the very short MD run-time provided per conformational sequence. Therefore, besides noting the dimer crossing angle orientation, the final angle is not taken into consideration.

7.2.3 United atom simulation configuration

MD simulations were performed on low complexity scaffold dimers AZ2, AZ2L6S and AZ2L10S.

7.2.3.1 Force field parameters

United atom force field parameters were set to those previously defined in section 4.2.1.

7.2.3.2 Simulation parameters

The temperature was set to 298 K to match the laboratory conditions of the assays presented in chapter 5. All other parameters are identical to those previously defined in section 4.2.2.

7.2.3.3 POPC bilayer construction

A POPC bilayer was constructed as per section 4.2.4, although, as umbrella sampling was not performed, the bilayer was reduced to 200 lipids and 11882 SPC water molecules.

7.2.3.4 Generation of initial configurations of the AZ2, AZ2L6S and AZ2L10S dimers

Initial configurations of AZ2, AZ2L6S and AZ2L10S were generated by subjecting each sequence to a global conformation search as defined in section 7.2.2. With a pK_a of 12.48, arginine was left charged in the presence of a membrane bilayer (255).

7.2.3.5 Insertion of dimers into a POPC bilayer

Dimers were inserted into pre-equilibrated POPC bilayers as previously defined in section 4.2.5.

7.2.3.6 Equilibration procedure

Dimers in POPC bilayers were equilibrated as previously explained in section 4.2.6, although each equilibration step was reduced to 5 ns, totalling 20 ns of equilibration time. The area-per-lipid of

the POPC bilayer was monitored during the equilibration process. Analysis was performed over a 40 ns unrestrained production runs using the *NPT* ensemble.

Heptad template	N-terminus	f	g	a	b	c	d	e	f	g	a	b	c	d	e	f	g	a	C-terminus
L17	GSS	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	TSR
L17(GG4)	GSS	L	L	L	L	L	L	G	L	L	L	G	L	L	L	L	L	L	TSR
L17L9Q	GSS	L	L	L	L	L	L	L	L	Q	L	L	L	L	L	L	L	L	TSR
AZ2	GSS	A	L	L	A	A	L	L	A	L	L	A	A	L	L	A	L	L	TSR
AZ2L6S	GSS	A	L	L	A	A	S	L	A	L	L	A	A	L	L	A	L	L	TSR
AZ2L10S	GSS	A	L	L	A	A	L	L	A	L	S	A	A	L	L	A	L	L	TSR
AZ2(GG4)$_L$	GSS	A	L	L	A	A	G	L	A	L	G	A	A	L	L	A	L	L	TSR
AZ2(GG4)$_A$	GSS	A	L	L	A	A	L	L	G	L	L	A	G	L	L	A	L	L	TSR

Table 7.1: TM domain sequences of low complexity scaffolds with additional motifs. The sequences are aligned to the heptad repeat schematic *a b c d e f g*, where red denotes the primary interface for the scaffold, and blue represents the secondary binding interface.

7.3 Results

7.3.1 CG low complexity scaffold free energy of self-association

7.3.1.1 L17

Umbrella sampling of L17 revealed a global free energy minimum of approximately -16.75 kJ mol^{-1} located 0.95 nm along the *inter*helical reaction coordinate (Figure 7.1). Histogram analysis and profiles built from 0.1 μs increments of data points (Figure 7.2) demonstrate a consistent overlap between neighbouring histograms and a converging PMF profile. The final PMF profile (0.1 μs–1 μs) resulted in a global free energy minimum which failed to overlap previous additive time series PMF profiles, yet the reported error would suggest that further sampling may not lead to a significant adjustment to the free energy minimum. The smaller histograms seen in Figure 7.2 (A), were additional umbrella windows of 200 ns with a harmonic potential of 800 kJ mol^{-1}. They were used to ensure adequate coverage of the reaction coordinate and overlap between neighbouring histograms. It turns out, however, that the initial 1 μs umbrella window simulations provided adequate sampling without the additional histogram windows. They have been included in the final PMF profile even though their absence is indifferent to the final result. The difference of 1.7 nm of umbrella windows along the reaction coordinate between PMF profile and histogram analysis was not included in the final PMF output of Figure 7.1 as the PMF profile began to converge within 2.3 nm. The trajectory that best matched the location of the global free energy minimum returned two clusters. Of the trajectory, 95% of the recorded time-frames were within the 0.5 Å RMSD cluster cut off (Figure 7.3 (A)). Only one other cluster was identified although the number of matching frames only constituted approximately 0.87% of the total trajectory (Figure 7.3 (B)). Both clusters were made from structures with a left-handed helix-helix crossing orientation.

A conformational atomistic search using CHI returned six clusters (Figure 7.4). The averaged structures A-F clearly show a series of left-handed helix-helix crossing orientations with leucine side chains stacking one above the other along the length of the helix-helix interface. Although a polyleucine sequence would suggest non-specific binding, experimental reports indicate that a polyleucine sequence adopts a small left-handed helix-helix crossing angle due to conformational restrictions from leucine side chain rotomeric entropic restrictions (54).

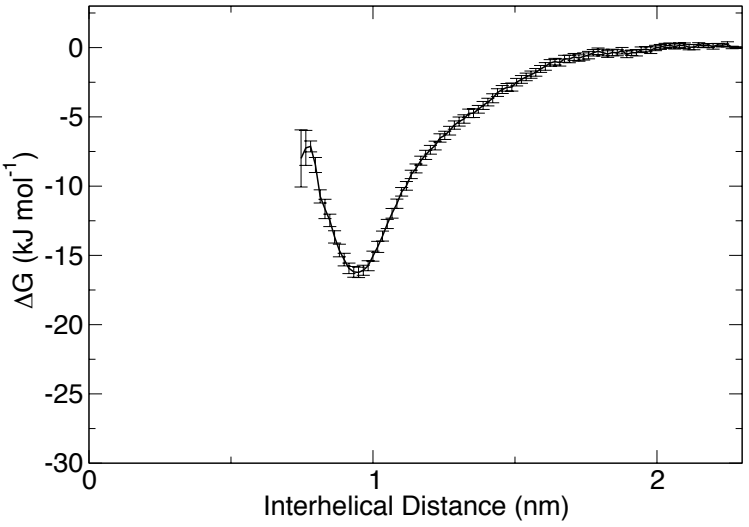

Figure 7.1: L17 PMF profile of self-association. Error was estimated over 200 bootstrap iterations. The global free energy minimum is estimated to be -16.75 kJ mol^{-1}, located 0.95 nm along the *inter*helical reaction coordinate.

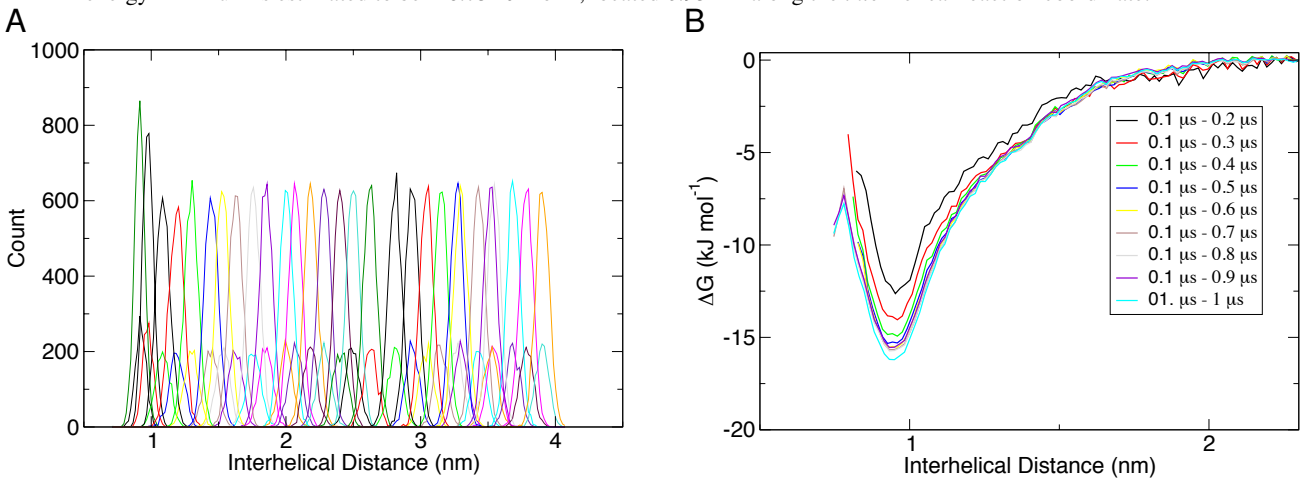

Figure 7.2: (A) L17 umbrella sampling histogram coverage along the *inter*helical reaction coordinate. (B) L17 converging PMF profiles from 0.1 μs to 1 μs in 0.1 μs increments.

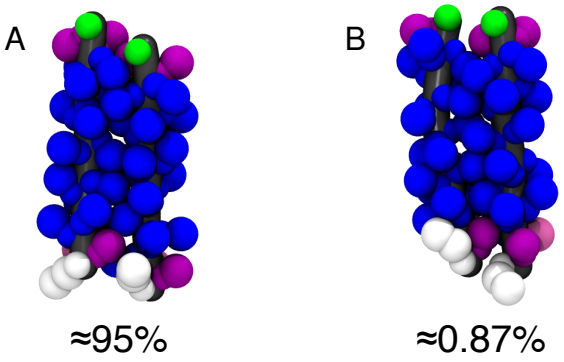

Figure 7.3: Each representative structure (A and B) from clusters with more than 10 structures from the CG trajectory pertaining to lowest free energy of L17 form a left-handed helix-helix crossing orientation. Glycine green, leucine blue, serine purple, tyrosine mauve, and arginine white.

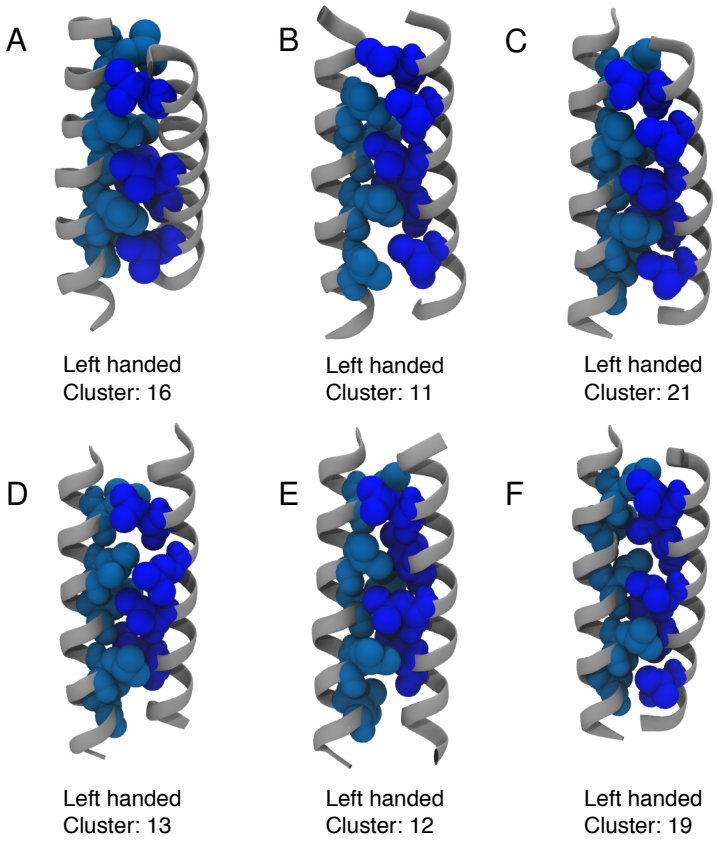

Figure 7.4: Average structures from the six identified clusters from an atomistic conformational search of L17 using CHI. Averaged structures A-F are left handed helix-helix dimers. For clarity, only the leucines at the helix-helix interface are shown in blue space filling sphere.

The CHI interaction energies of each averaged structure are presented in Figure 7.5 along with the total interaction interaction energy value. Interaction energy profile F presents the structure with the highest interaction energetics contributed from packed side chains. The helix-helix energetic contributions arise from the interacting side chains xxxxxLxxxLxxLxxxLxxL. This sequence matches a repeat of ‘a’ and ‘d’ sites on a coiled coil. Further more, the respective average structure has a left handed dimer, indicative of most heptad repeat mediated dimers.

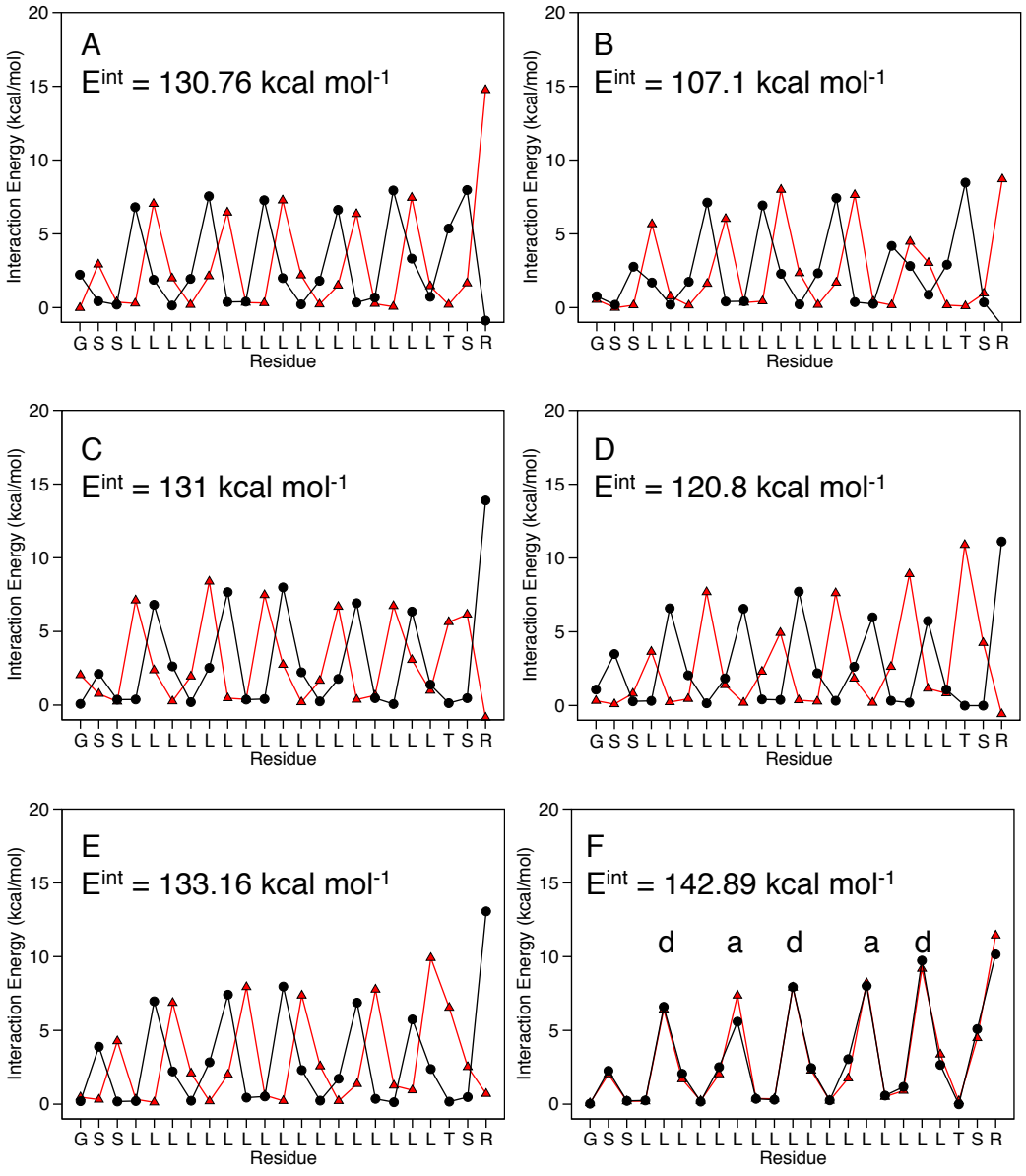

Figure 7.5: Atomistic L17 Interaction energies for matching structures in Figure 7.4. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

7.3.1.2 L17(GG4)

Umbrella sampling of L17(GG4) revealed a single free energy minimum of approximately -19.18 kJ mol[1], located 0.85 nm along the *inter*helical reaction coordinate (Figure 7.6). Histogram analysis (Figure 7.7 (A)) demonstrated significantly overlapping distributions across the entire reaction coordinate and profiles built from 0.1 μs increments of data points (Figure 7.7 (B)) indicate converged sampling. CG clusters isolated from the trajectory matching the location of the global free energy minimum using a 0.5 Å RMS cut-off yielded right-handed helix-helix crossing orientations (Figure 7.8 (A), (B), (B), and (E)) with the exception to one left-handed dimer (Figure 7.8 (D)). In all but the left-handed structure, the GG4 interaction motif was found at least partially, if not completely, buried at the helix-helix interface. The glycine residues in the right-handed cluster are orientated out, towards the bilayer lipid chains.

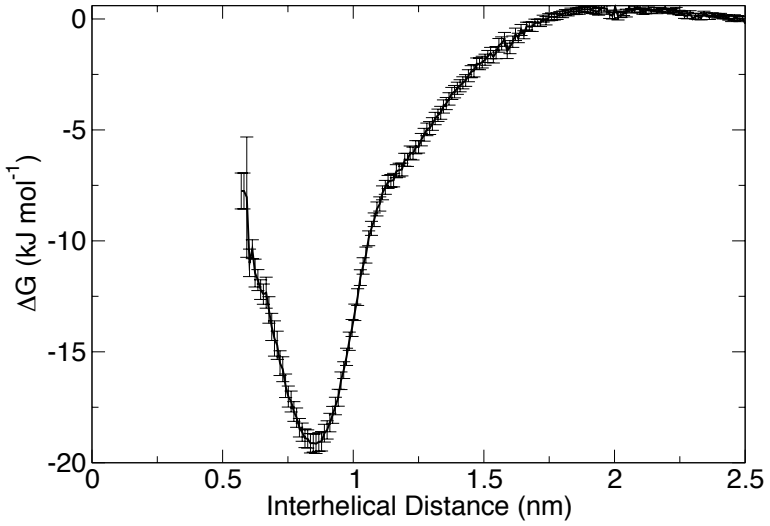

Figure 7.6: L17(GG4) PMF profile of self-association. Error was estimated over 200 bootstrap iterations. The global free energy minimum is estimated to be -19.18 kJ mol^{-1}, located 0.85 nm along the *inter*helical reaction coordinate.

A conformational atomistic search using the CHI algorithm returned five clusters, each containing at least ten structures (Figure 7.9). An average structure from each cluster revealed that none of the clusters were driven by the GG4 interaction motif, rather favouring a leucine side chain packed arrangement. All five structures adopted a left-handed helix-helix crossing orientation, indicative of a left handed heptad repeat. The CHI conformations do not compare well with the CG representative structures found at the free energy minimum. CG structures demonstrate a propensity for packing the GG4 motif into the helix-helix interface, whilst CHI conformations

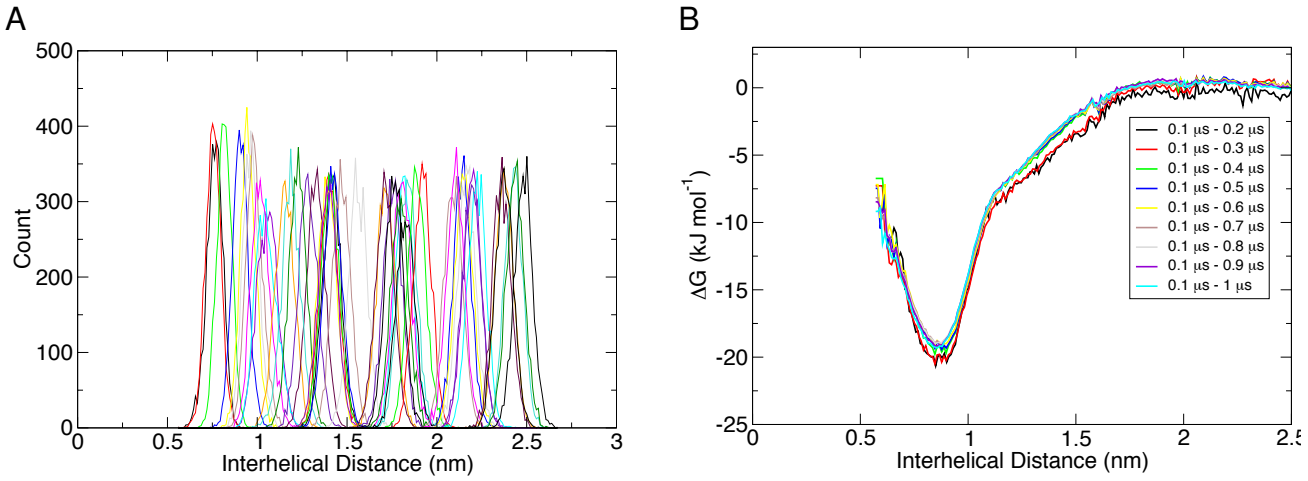

Figure 7.7: (A) L17(GG4) umbrella sampling histogram coverage of the *inter*helical reaction coordinate over 1 μs. (B) L17(GG4) PMF profiles built from 0.1 μs increments of data points.

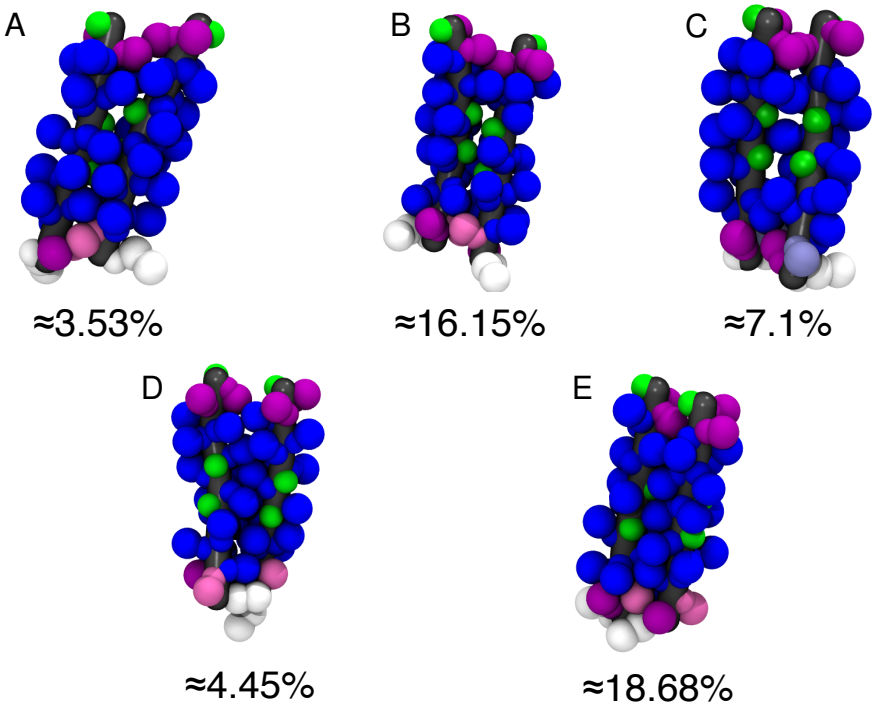

Figure 7.8: Representative of L17(GG4) dimers from cluster analysis at the global free energy minimum found approximately 0.85 nm along the *inter*helical reaction coordinate. (A) Right helix-helix cross, (B) right helix-helix cross, (C) right helix-helix cross, (D) left helix-helix cross, and (E) right helix-helix cross. Glycine green, leucine blue, serine purple, tyrosine mauve, and arginine white.

adopt a left handed orientation. Interaction energy profile E in Figure 7.10 yields the highest total interaction energy, with both helices packing using an 'a' and 'd' repeat pattern. Only one GG4 motif marginally contributed to helix-helix packing. One could hypothesis that the a GG4 mediated helix-helix interaction is favoured in an explicit bilayer due to the increase in the number of degrees of freedom leucine side chains has over glycine when orientated into the bilayer core.

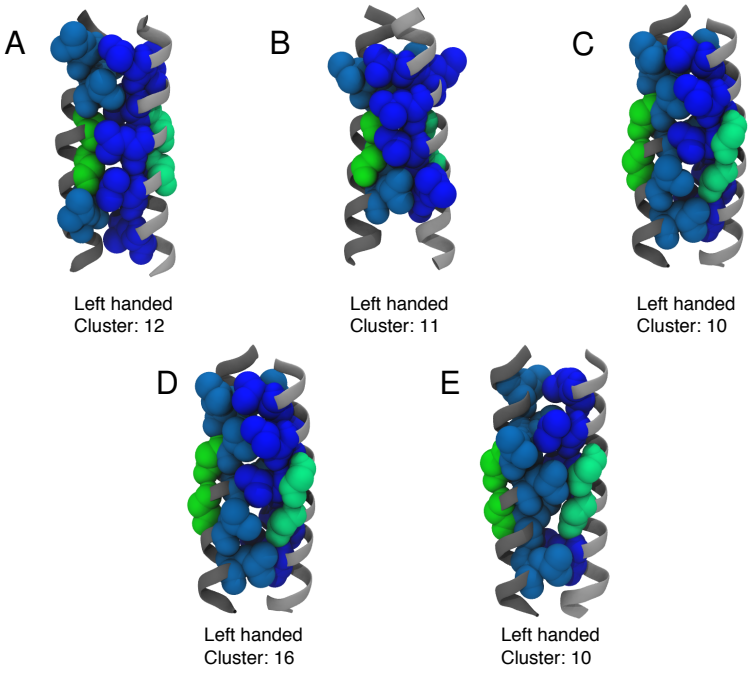

Figure 7.9: Five average structures from from an atomistic conformational search of L17(GG4) using CHI. All structures A-E were found to adopt a left helix-helix dimer. For clarity, only the leucines (blue) at the helix-helix interfaces are are shown in space filling sphere in addition to glycines (green) of the GG4 motifs.

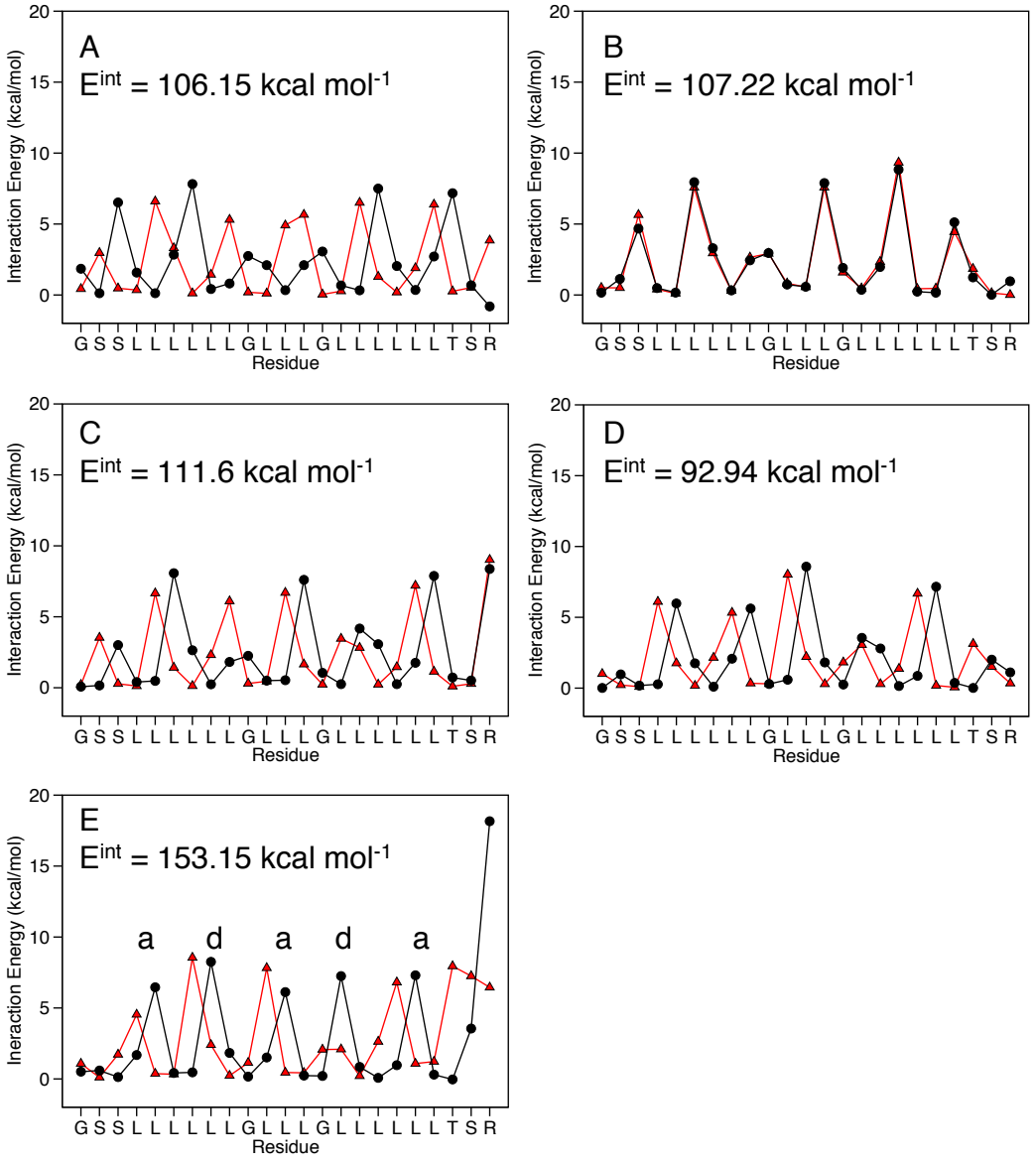

Figure 7.10: Atomistic L17GG4 Interaction energies for matching structures in Figure 7.9. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

7.3.1.3 L17L9Q

Umbrella sampling of L17L9Q dimers revealed two statistically similar free energy minima located approximately 0.5 nm and 0.95 nm along the reaction coordinate with a value of -10.5 kJ mol^{-1} (Figure 7.11). There is a shallow local minimum located approximately 1.5 nm along the reaction coordinate with a free energy value of -8.75 kJ mol^{-1}. Due to insufficient sampling towards the later half of the reaction coordinate, additional umbrella windows with umbrella restraints of 800 kJ mol^{-1} were set to run for 200 ns (see Figure 7.12 (A)). PMF profiles from 0.1 μs increments of data points (Figure 7.12 (B)) suggest that the PMF profile began to converge after 0.7 μs.

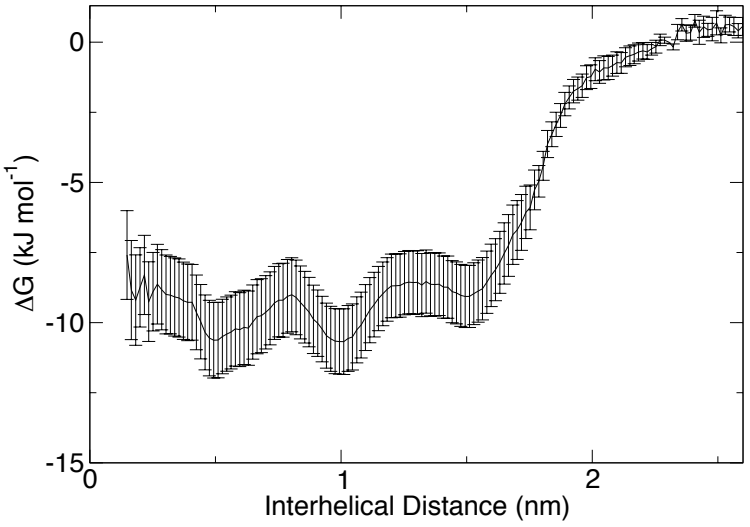

Figure 7.11: L17L9Q PMF profile of self-association. Error was estimated over 200 bootstrap iterations. The global free energy minimum is estimated to be -10.5 kJ mol^{-1}, located 0.95 nm along the *inter*helical reaction coordinate. There are in total three energy minima.

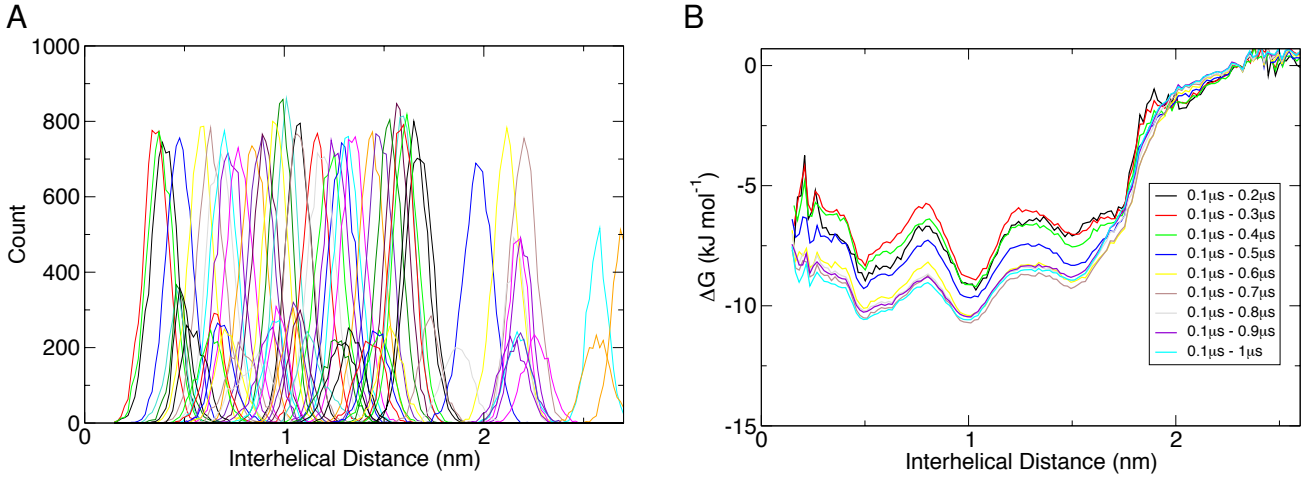

Figure 7.12: (A) L17L9Q umbrella sampling histogram coverage of the *inter*helical reaction coordinate. (B) L17L9Q converging PMF profiles.

Trajectories from both statistically equivalent free energy minima (0.5 nm and 0.95 nm) were subjected to cluster analysis with an RMSD cut-off of 0.5 Å. The trajectory pertaining to the free energy minimum located at 0.95 nm, returned four clusters (Figure 7.13), all of which were right-handed dimers with both glutamine side-chains orientated into the bilayer lipid chains. The leucine packed helix-helix interface accompanied by the location of the free energy minimum as both indicative to the propensity for self-association of L17. However, the conformations of L17L9Q suggest less favourable right-handed packing and an energetic penalty of orientating glutamine into the POPC hydrophobic core.

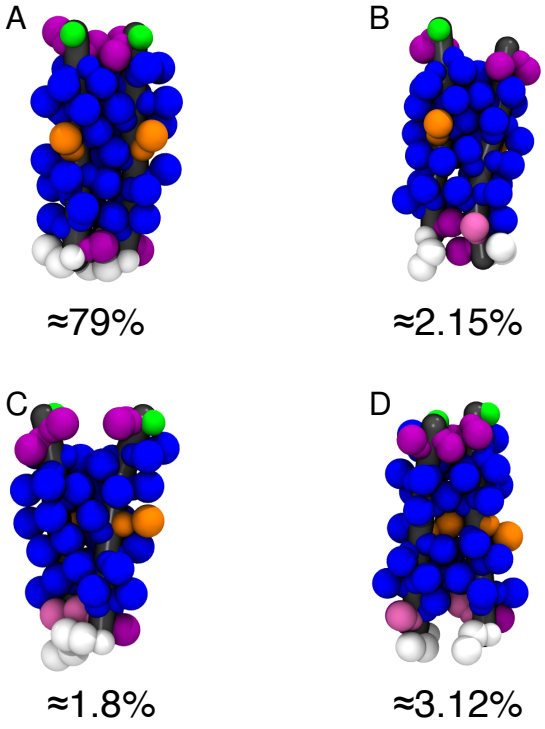

Figure 7.13: Representative from each of the four clusters (A-D) with more than 10 structures from the CG trajectory pertaining to free energy minimum at 0.95 nm along the *inter*helical reaction coordinate of L17L9Q. All four identified clusters are constructed from right helix-helix dimers. Glycine green, leucine blue, serine purple, tyrosine mauve, glutamine orange, and arginine white.

The trajectory pertaining to the free energy minimum located at 0.5 nm, returned ten clusters of which Figure 7.14 (A), (B), (G) and (I) formed left-handed dimers, (C), (E), (F) and (J) formed right-handed dimers, and (D) and (H) had near to no crossing angle (zero). All structures suggested a clear interaction between glutamine side chains.

Initial cluster analysis from a CHI conformational search failed to return any clusters from an *inter*helical distance of 1 nm. After adjusting the distance to 1.2 nm, eleven clusters were returned

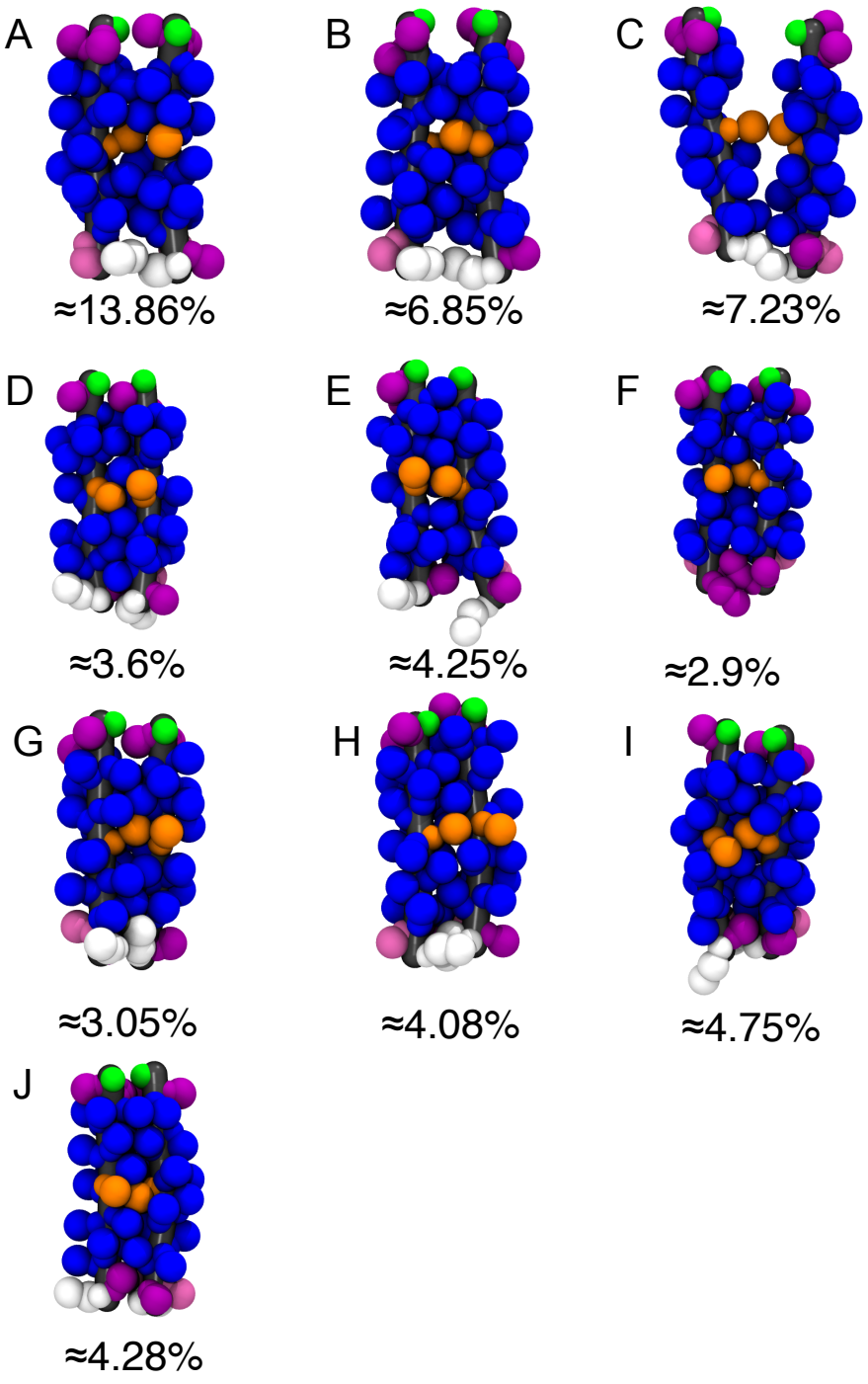

Figure 7.14: Representatives from clusters with more than 10 structures from the CG trajectory pertaining to free energy minimum at 0.5 nm along the *inter*helical reaction coordinate of L17L9Q. (A) left helix-helix cross, (B) left helix-helix cross, (C) right helix-helix cross, (D) zero helix-helix cross, (E) right helix-helix cross, (F) right helix-helix cross, (G) zero helix-helix cross, (H) left helix-helix cross, (I) left helix-helix cross, and (J) right helix-helix cross. Glycine green, leucine blue, serine purple, tyrosine mauve, glutamine orange, and arginine white.

with at least ten structures per cluster. Of the eleven averaged structures (Figures 7.15), nine structures adopted a left-handed helix-helix crossing orientation and the remaining two structures formed right-handed dimers. There was no discernible correlation between the glutamine position

and crossing angle orientation between the CHI models and the CG cluster representations from the 0.5 nm free energy minimum.

The interaction energy profiles A in Figure 7.16 and I in Figure 7.17 from CHI clustering analysis of L17L9Q, shows a symmetrical packed helix-helix interface. Both helices interact with a clear ‘a’ and ‘d’ site contribution, and given that both structures adopt a left-handed helix-helix crossing orientation, the side chain packing is very similar to a heptad repeat structure. Interestingly, there is little to no interaction between glutamine residues, with perhaps a marginal interaction energy contribution from the alignment of side chain amide group with the neighbouring backbone. Finally, the third most significantly contributing set of interaction profiles (G) suggests that glutamine is contributing by approximately 5 kcal mol^{-1} and 12 kcal mol^{-1}. The matching structure would indicate a potential hydrogen bond between the side chain amide group and the opposite side chain oxygen.

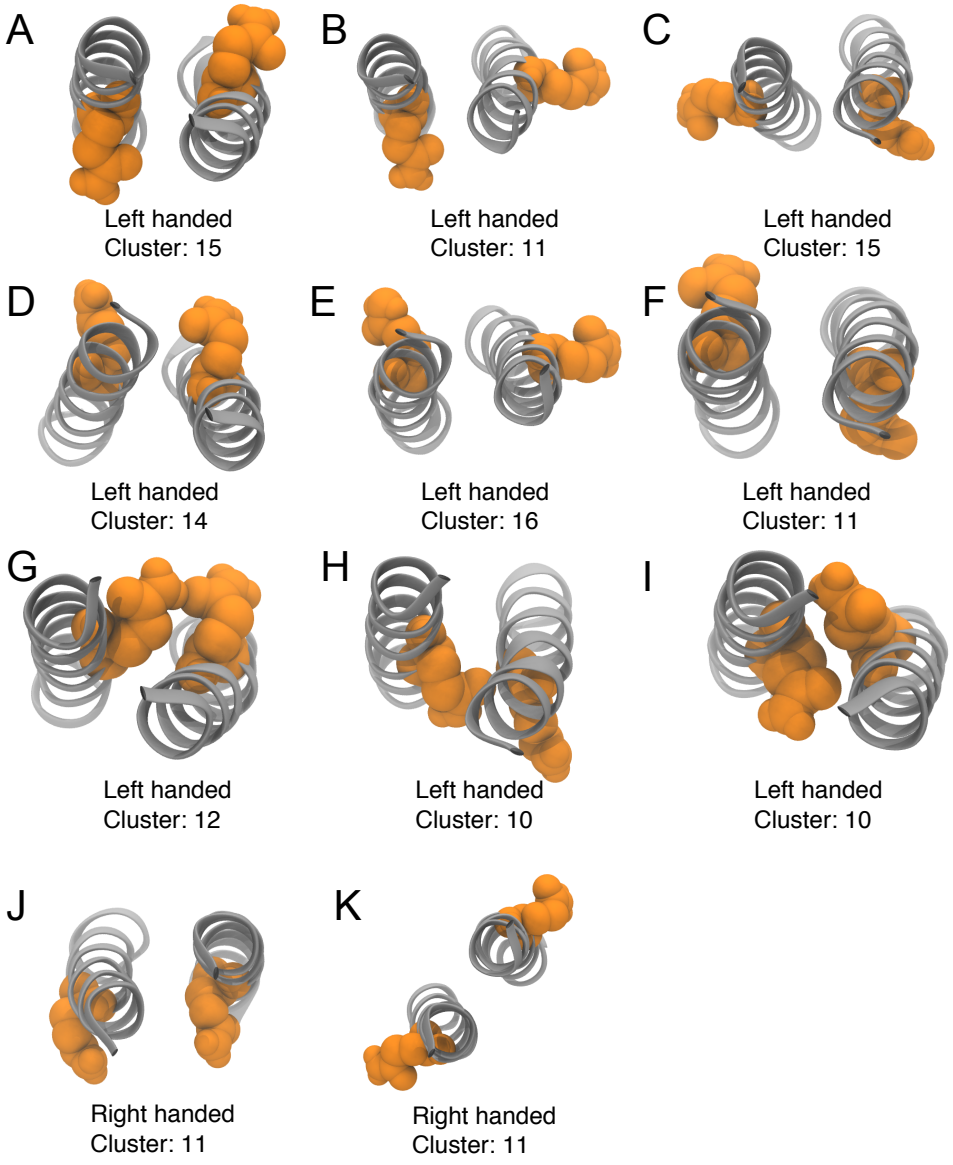

Figure 7.15: Eleven average structures from a conformational search of L17L9Q using CHI. Nine of the eleven averaged structures are left helix-helix dimers, structure A-I. Right handed dimers are structures J and K. For clarity, only the glutamine residues (orange) in space filling spheres are shown on a ribbon-helical backbone.

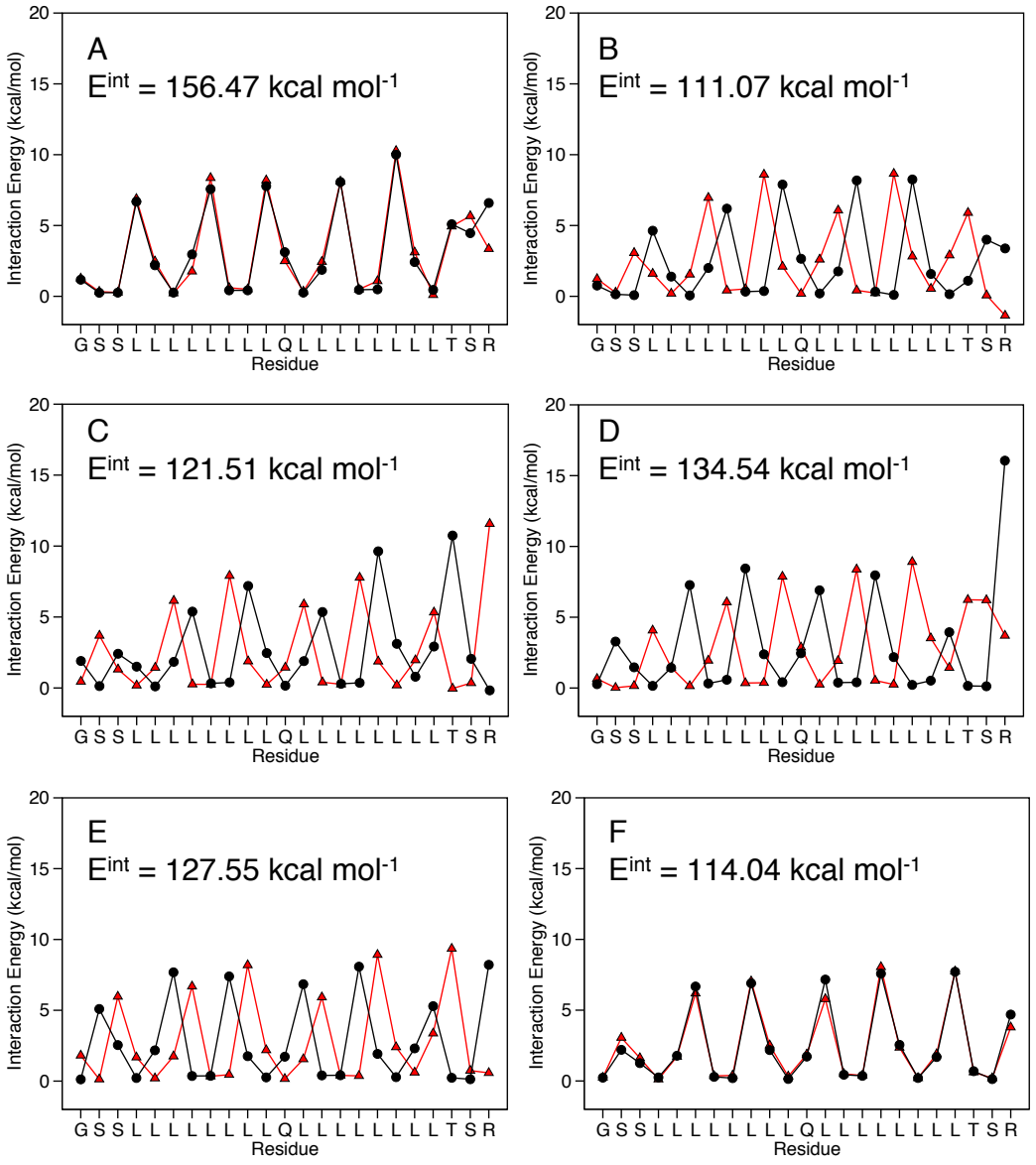

Figure 7.16: Atomistic L17L9Q Interaction energies for structures A-F in Figure 7.15. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

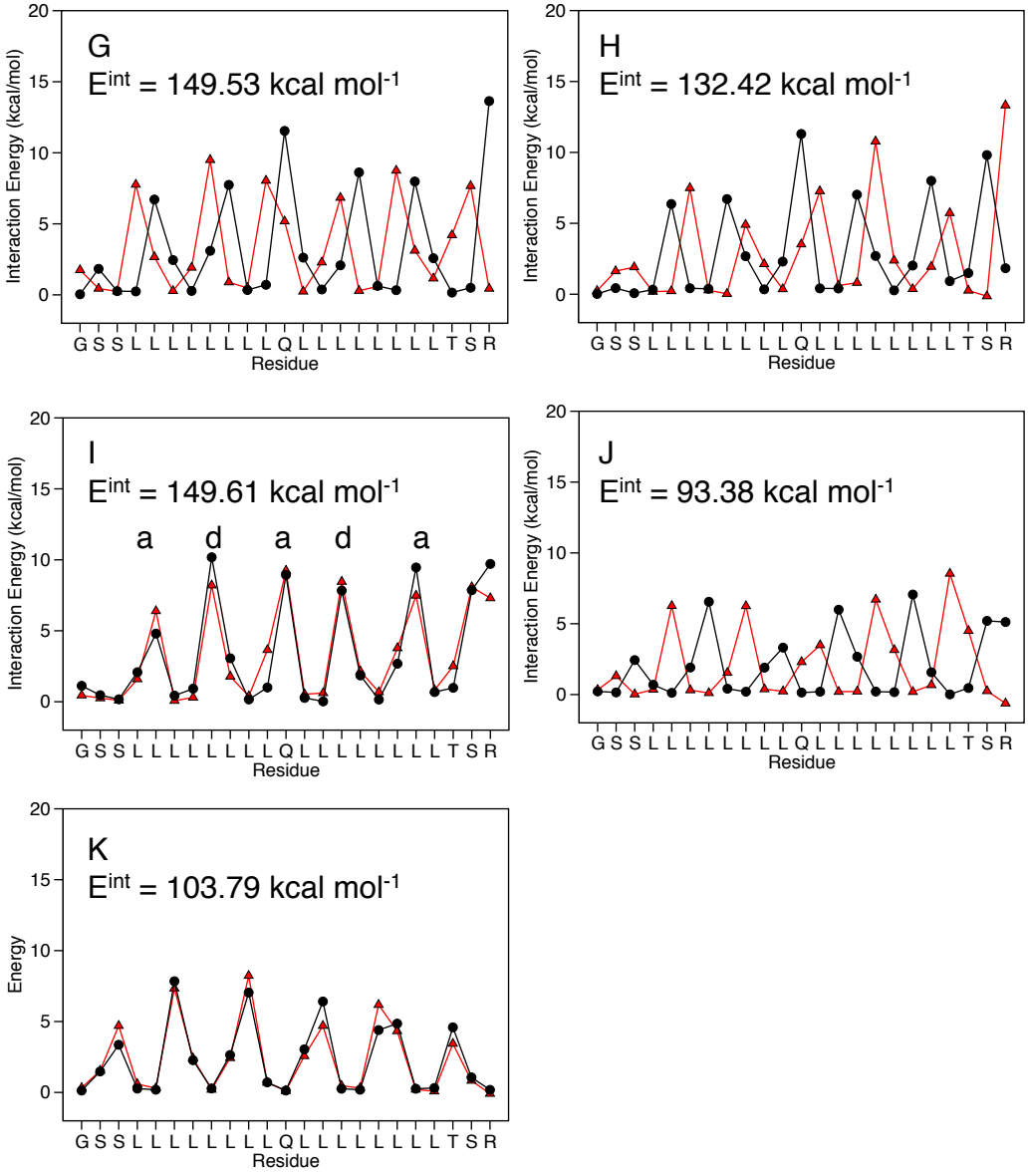

Figure 7.17: *Continued...* Atomistic L17L9Q Interaction energies for structures G-K in Figure 7.15. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

7.3.1.4 AZ2

Umbrella sampling of AZ2 revealed a deep and broad global free energy minimum, approximately 0.25 nm across (Figure 7.18). The trajectory for further analysis was taken at the minimum, approximately 1.2 nm along the reaction coordinate, with a ΔG value of -24.75 kJ mol^{-1}. Histogram analysis (Figure 7.19 (A)) demonstrated sufficient sampling along the reaction coordinate even though the data points to several histogram distributions within the energy well region were unevenly distributed. The PMF profiles built from 0.1 μs increments of data points (Figure 7.19 (B)) suggest that the location of the global free minimum at 1.2 nm had converged within the allotted simulation time.

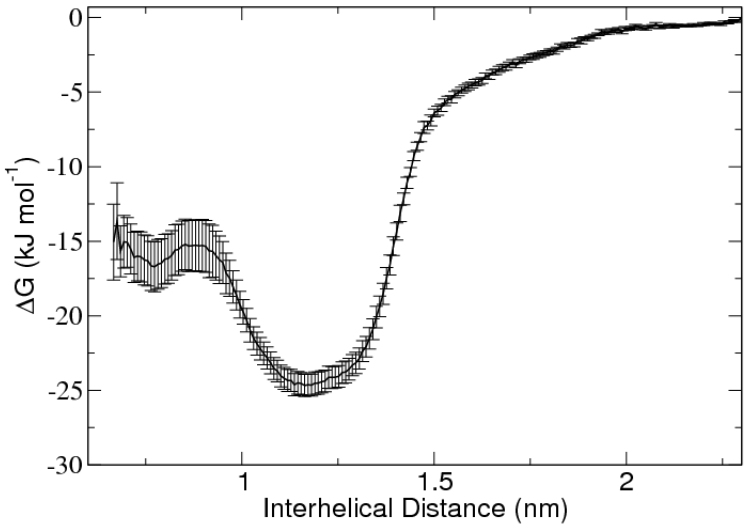

Figure 7.18: AZ2 PMF profile of self-association. Error was estimated over 200 bootstrap iterations. The global free energy minimum is estimated to be -24.75 kJ mol^{-1}, located 1.2 nm along the *inter*helical reaction coordinate.

The trajectory pertaining to the global free energy minimum returned two clusters constituting approximately 84% and 14% of the recorded trajectory (Figure 7.20 (A) and (B), respectively). Both clusters adopted a left-handed helix-helix crossing orientation with an alanine packed helix-helix interface.

A conformational search using the CHI algorithm returned eight clusters, each containing at least ten structures (Figure 7.21). An averaged structure from each cluster revealed all but one had adopted a left-handed helix-helix crossing orientation. Interestingly, the right-handed exception, structure G, had adopted an alanine packed helix-helix interface. This compares well with the established notion that the packing of small residues (serine, alanine, glycine) conforms to a right handed

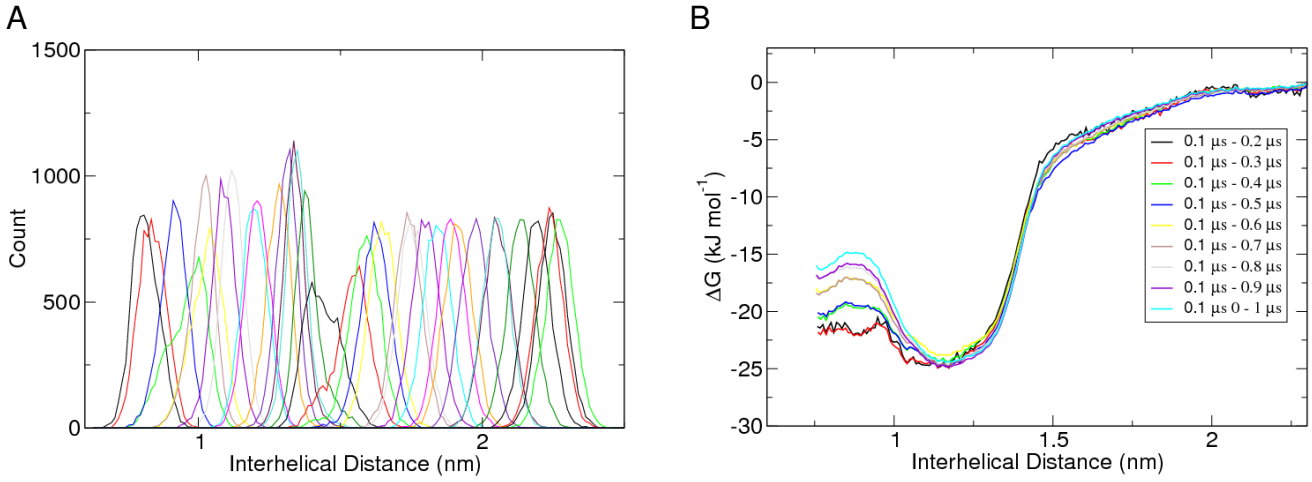

Figure 7.19: (A) AZ2 umbrella sampling histogram coverage of the *inter*helical reaction coordinate. (B) AZ2 converging PMF profiles.

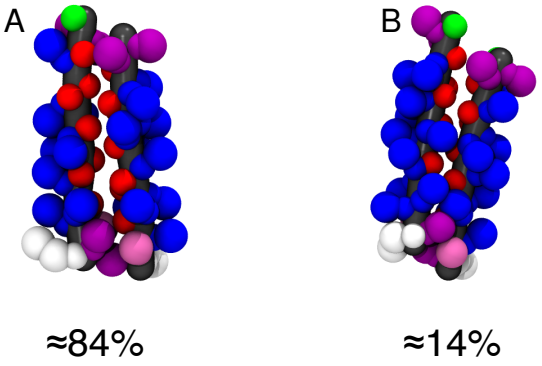

Figure 7.20: Representatives from clusters with more than 10 structures from the CG trajectory pertaining to free energy minimum at 1.2 nm along the *inter*helical reaction coordinate of AZ2. (A) left helix-helix cross, and (B) left helix-helix cross. Glycine green, leucine blue, serine purple, tyrosine mauve, alanine red, and arginine white.

dimer rather than the 'knobs-into-holes' left handed conformation applicable to the longer and more flexible side chain of leucine (224). Of the left-handed clusters, one averaged structure adopted an alanine packed helix-helix interface similar to the CG cluster analysis, whilst the remaining left-handed structures returned an alanine-leucine packed helix-helix interface. According to the total interaction energies presented in Figure 7.22, structure F yields the greatest number of side chain energetic contributions. Both helices are seen to adopt an 'a' and 'd' repeat interhelical packing motif with a left handed crossing orientation.

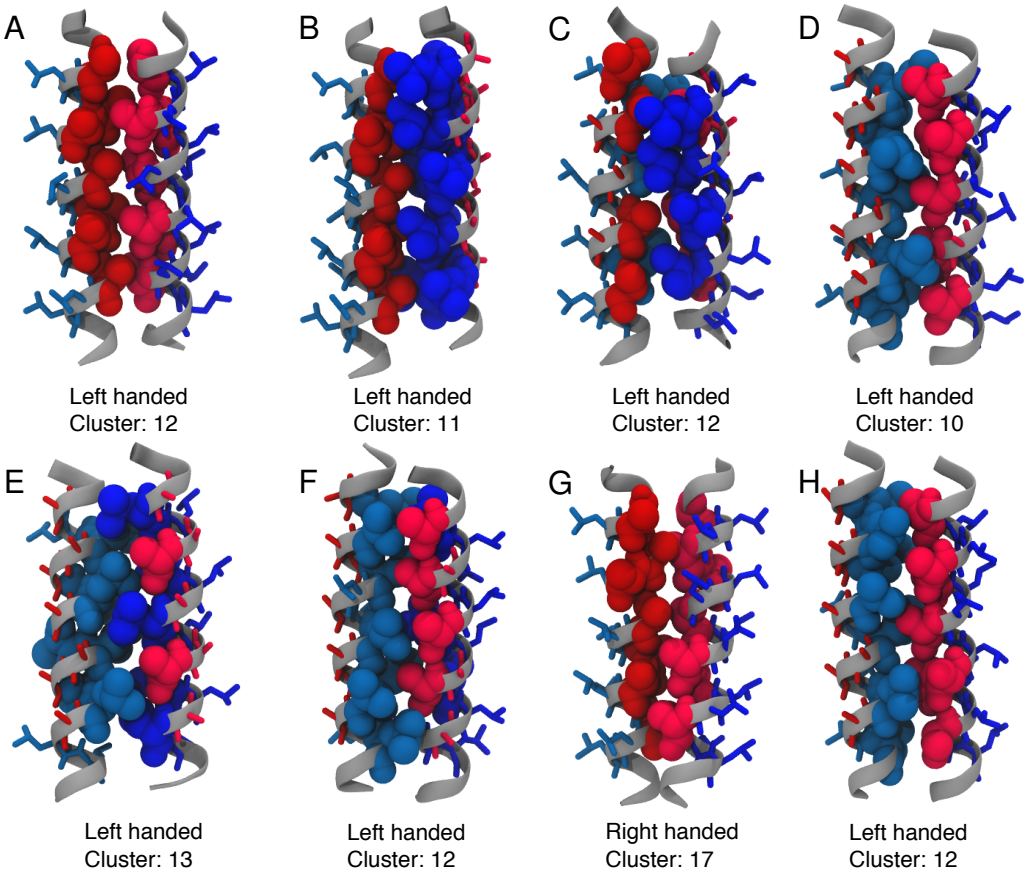

Figure 7.21: Average structures from eight AZ2 clusters. Seven of the eight averaged structures, A-F and H, are left-handed dimers. Structure G is a right handed dimer. For clarity, only residues present on the helix-helix interface are shown in space filling spheres, all other residues are represented in stick-form. Leucine is shown in blue, and alanine in red.

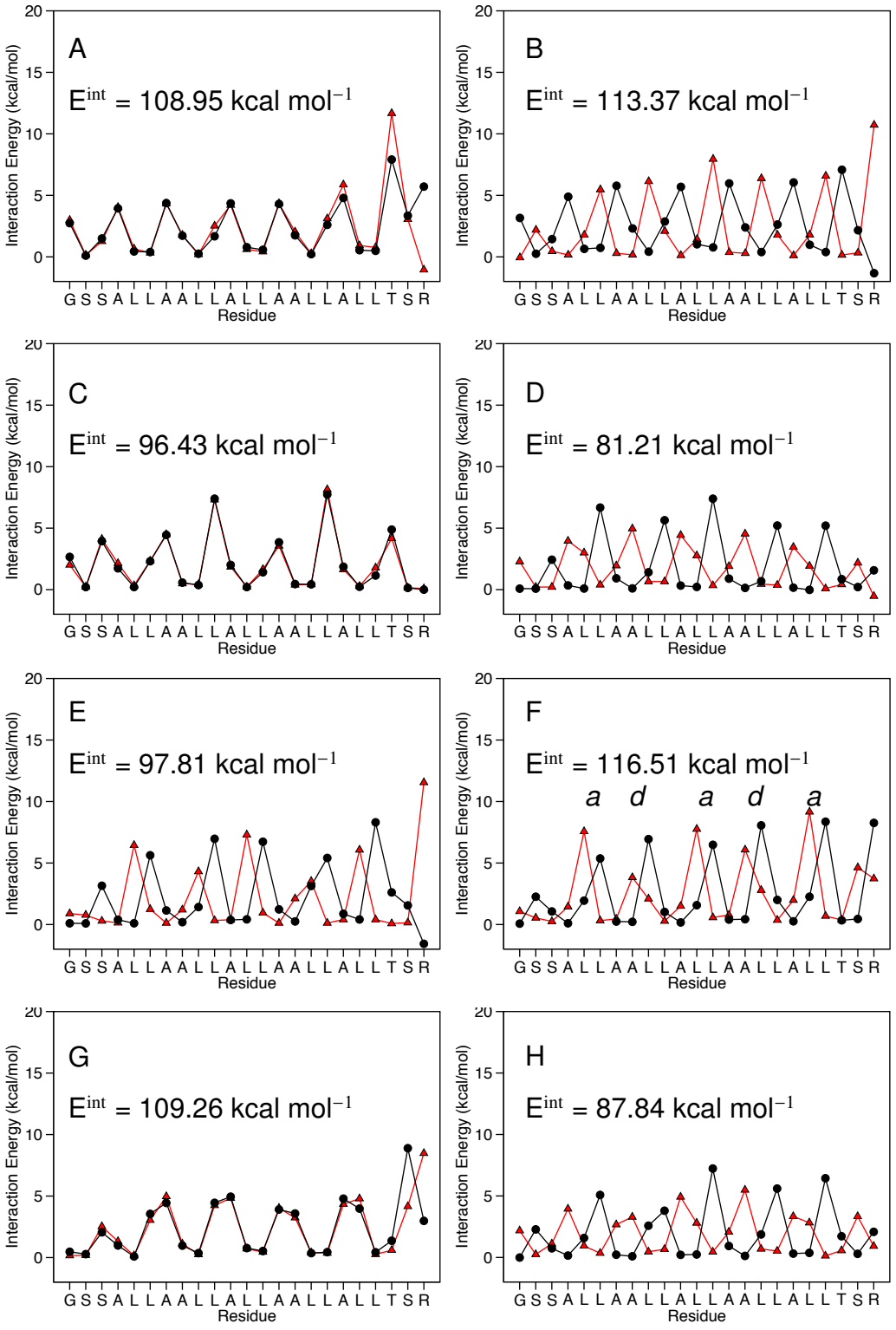

Figure 7.22: Atomistic AZ2 Interaction energies for structures A-H in Figure 7.21. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

7.3.1.5 AZ2(GG4)$_L$

Umbrella sampling of AZ2(GG4)$_L$ revealed a global free energy minimum with a value of -24 kJ mol^{-1}, located 0.7 nm along the *inter*helical reaction coordinate (Figure 7.23). It is interesting to note that the global free energy minimum is accompanied by a statistically equivalent region of the free energy, beginning from the start of the reaction coordinate. Further along the reaction coordinate, a local minimum with an approximate value of -21 kJ mol^{-1} was found at 1.05 nm. Histogram analysis (Figure 7.24 (A)) indicated sufficient overlap between umbrella histograms. The denser regions of histograms compare well with the location of free energy minima. The uneven distribution does not prevent the PMF profiles from converging as seen by generating profiles with 0.1 μs increments of data points (Figure 7.24 (B)).

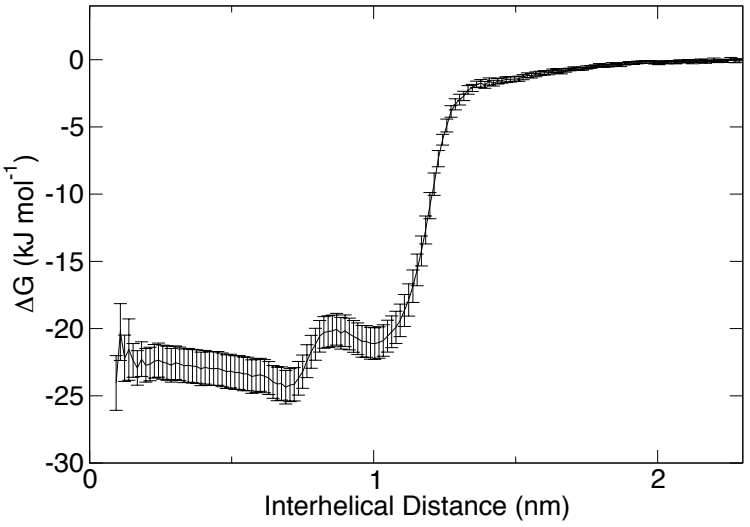

Figure 7.23: AZ2(GG4)$_L$ PMF profile of self-association. Error was estimated over 200 bootstrap iterations. The global free energy minimum is estimated to be -24 kJ mol^{-1}, located 0.7 nm along the *inter*helical reaction coordinate.

CG cluster isolated from the global free energy minimum trajectory returned two clusters. Approximately 94% of the recorded trajectory (Figure 7.25 (A)) adopted a right-handed helix-helix crossing orientation with a polyleucine helix-helix interface along with both GG4 interaction motifs buried at the interface. The second cluster (Figure 7.25 (B)), represents only 0.25% of the recorded trajectory but also adopted a right-handed helix-helix orientation.

Cluster analysis performed on the second free energy minimum, approximately 1 nm along the reaction coordinate, yielded four clusters (Figure 7.26). Approximately 20.03% of the trajectory adopted a left-handed helix-helix crossing orientation (Figure 7.26 (A)), with alanine residues

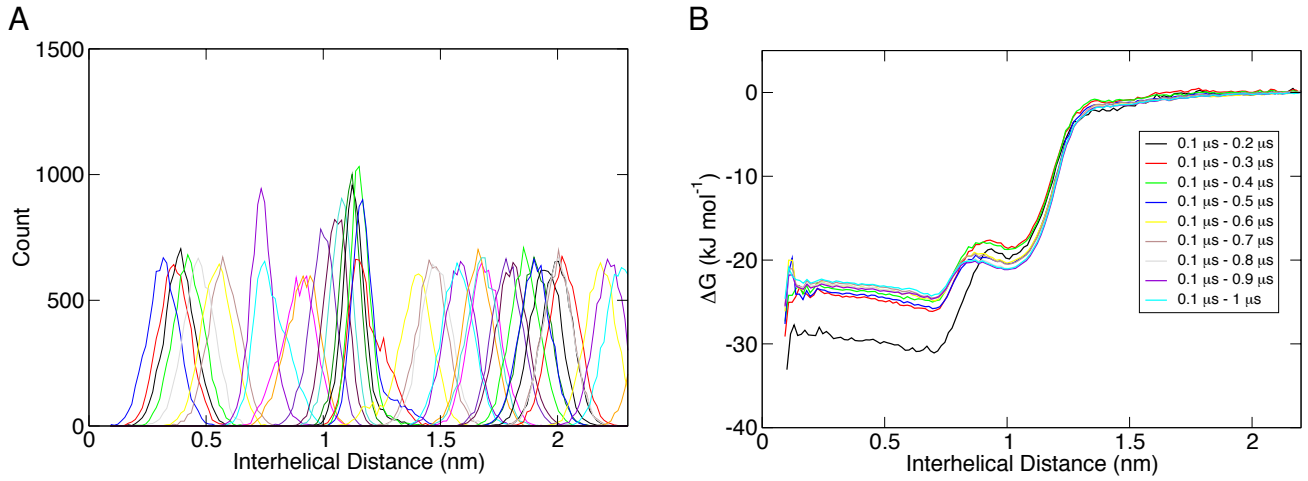

Figure 7.24: (A) AZ2(GG4)$_L$ umbrella sampling histogram coverage of the *inter*helical reaction coordinate. (B) AZ2(GG4)$_L$ converging PMF profiles from 0.1 μs to 1 μs in 0.1 μs increments.

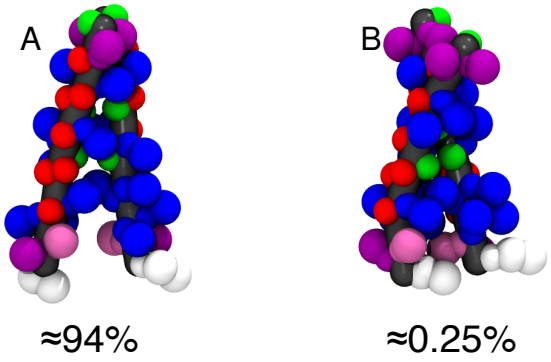

Figure 7.25: Representatives from clusters with more than 10 structures from the CG trajectory pertaining to free energy minimum at 0.7 nm along the *inter*helical reaction coordinate of AZ2(GG4)$_L$. (A) right helix-helix cross, and (B) right helix-helix cross.Glycine green, leucine blue, serine purple, tyrosine mauve, alanine red, and arginine white.

supporting dimerisation toward the C terminus. The structure in Figure 7.26 (B) suggests that approximately 6.75% of trajectory adopted a zero-handed helix-helix orientation, with leucine packing at the helix-helix interface. The remaining two representative structures, Figures 7.26 (C) and (D), represent clusters of right-handed helix-helix crossing orientations. Both are shown to have leucine residues packing at the helix-helix interface.

A conformational search using the CHI algorithm returned seven clusters each containing at least ten structures (Figure 7.27). Of the seven, five averaged structures adopted left-handed crossing orientation (A-E) whilst the remaining two (F, and G) yielded right-handed helix-helix crossing orientations. Two of the left-handed conformations had an alanine packed helix-helix interface (C, and D) whilst the remaining dimers adopted a polyalanine-polyleucine helix-helix interface (A, B, and E). A single GG4 motif was present in all but structures C and D.

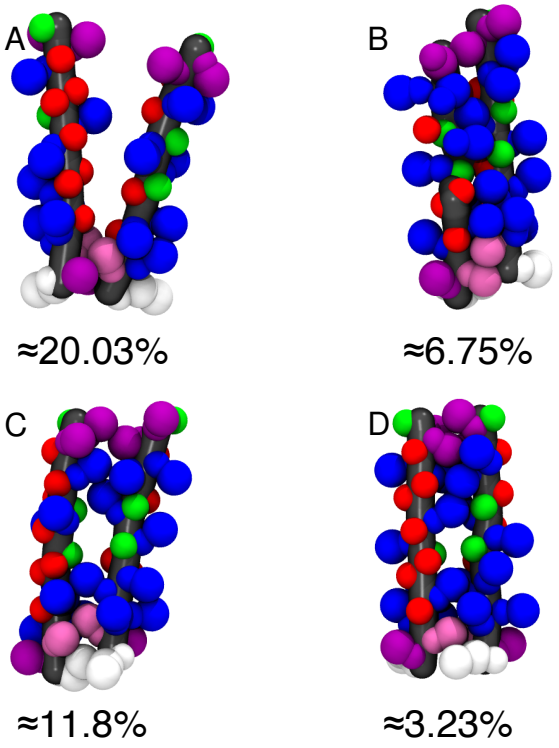

Figure 7.26: Representative of AZ2(GG4)$_L$ dimers from cluster analysis at the free energy minimum found approximately 1 nm along the *inter*helical reaction coordinate. (A) left helix-helix crossing, (B) zero helix-helix crossing, (C) right-helix-helix crossing, and (D) right helix-helix crossing. Glycine green, leucine blue, serine purple, tyrosine mauve, alanine red, and arginine white.

The interaction energy plots relating to the averaged structures from CHI cluster analysis are given in Figure 7.28. Interaction energy plots B-E yield the highest total interaction energy, the discernible difference in total energy being attributed to the final three polar residues, tyrosine, serine and arganine. Structure B demonstrates a helix-helix packing contribution of sites ‘a’ and ‘d’ along both helices, although the side chains responsible are not symmetrical given that the average structure demonstrates leucines packing against alanines with one GG4 motif in between. Interaction energy profiles C and D demonstrate a symmetrical packing of alanine side-chains, conforming to a ‘a’ and ‘d’ repeat pattern. Finally, structure E yields an *inter*helical packing arrangement of alanine with leucine similar to the structure B.

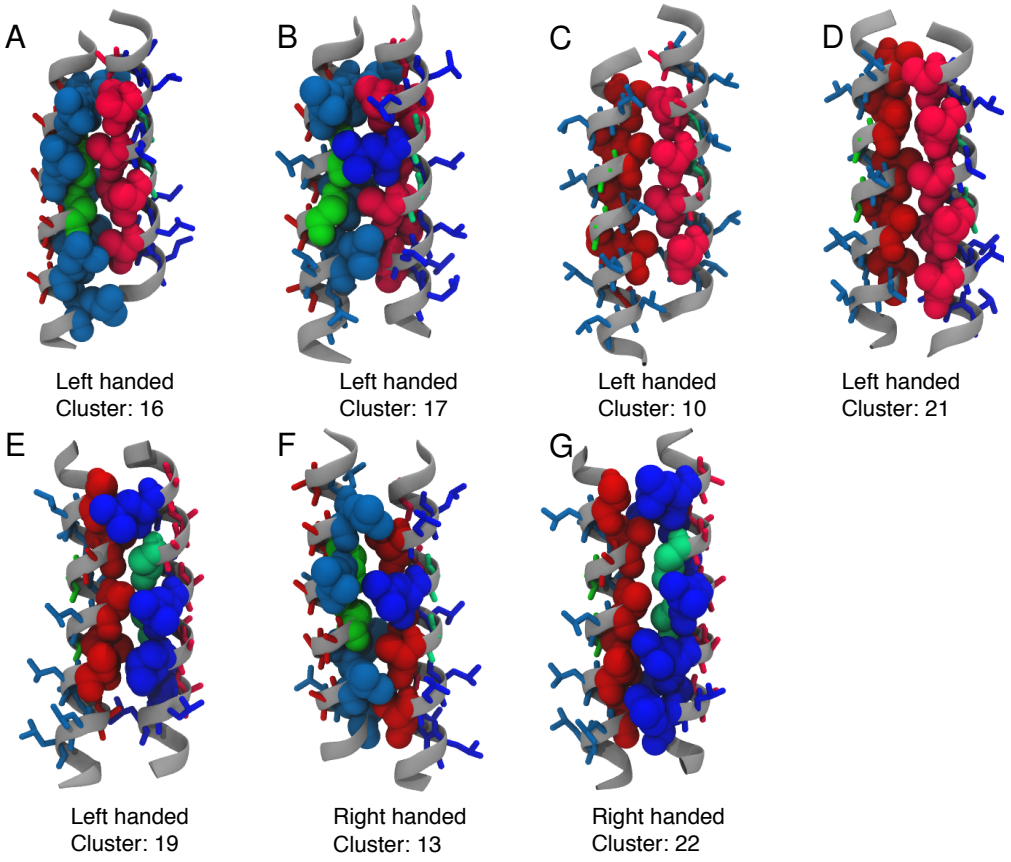

Figure 7.27: Average structures from seven AZ2(GG4)$_L$ clusters. Averaged structures A-D are left-handed dimers. Structures E and F are right handed dimers. For clarity, only residues present on the helix-helix interface are shown in space filling spheres. All other residues are represented in stick-form. Leucine is shown in blue, alanine in red, and glycine in green.

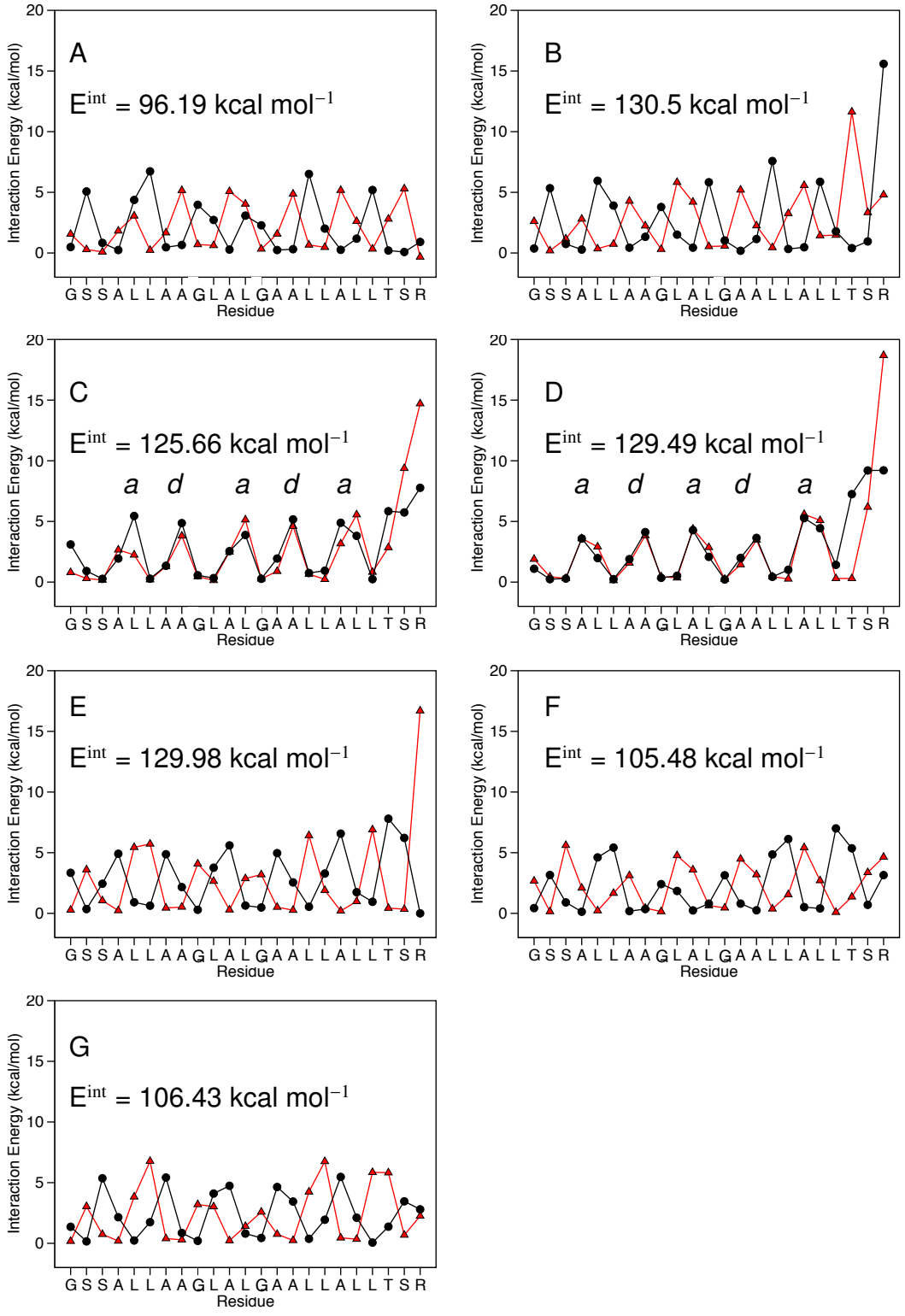

Figure 7.28: Atomistic AZ2(GG4)$_L$ Interaction energies for structures A-G in Figure 7.27. For clarity, of the four profiles with the highest total interaction energy (B-E), only profiles C and D have been labeled with heptad repeat sites due to their symmetrical interactions. Profiles B and E conform to a heptad repeat interaction pattern, however, the interactions are not symmetrical. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

7.3.1.6 AZ2(GG4)$_A$

The global free energy minimum was approximately -27.4 kJ mol^{-1}, 0.75 nm along the *inter*helical reaction coordinate (see Figure 7.29). The location of the free energy minimum would suggest a close *inter*helical distance with respect to the *inter*helical distance between GG4 found at the interface of the AZ2(GG4)$_L$ dimer. The estimated error from bayesian bootstrap analysis suggests that the reaction coordinate had been sufficiently sampled. Data points taken from histograms between 0.1μs to 0.4μs failed to produce cohesive PMF profiles as depicted in Figure 7.30 (B), however, the time incremental PMF profiles rapidly converged after 0.5μs. The uneven appearance of umbrella histogram windows in Figure 7.30 (A) was due to an uneven distribution of data point. However, this did not affect the final PMF profile (0.1 μs - 1 μs of data points) given the sufficient overlap of histograms.

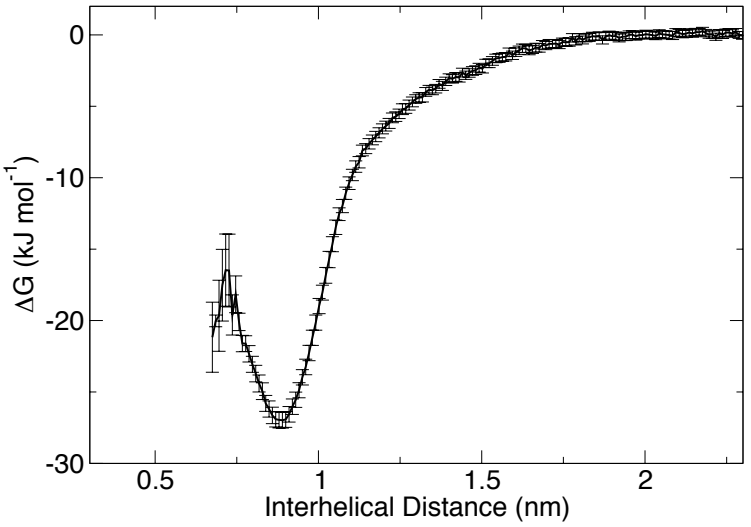

Figure 7.29: AZ2(GG4)$_A$ PMF profile of self-association. Error was estimated over 200 bootstrap iterations. The global free energy minimum is estimated to be -27.4 kJ mol^{-1}, located 0.75 nm along the *inter*helical reaction coordinate.

Cluster analysis performed over 0.1 μs to 1 μs of the trajectory from the umbrella window best matching the global free energy minimum returned three clusters. A structural representation of each cluster is depicted in Figure 7.31. Approximately 48.8 % of the trajectory adopted a right handed dimer, followed by 9.45% of the trajectory adopting a zero-crossing angle orientation. Finally, a further 2.6% of the trajectory adopted a right-handed dimer. The GG4 motif of all three cluster representations were found buried at the helix-helix interface.

A CHI conformational search of an AZ2(GG4)$_A$ dimer, returned eight clusters with more than or

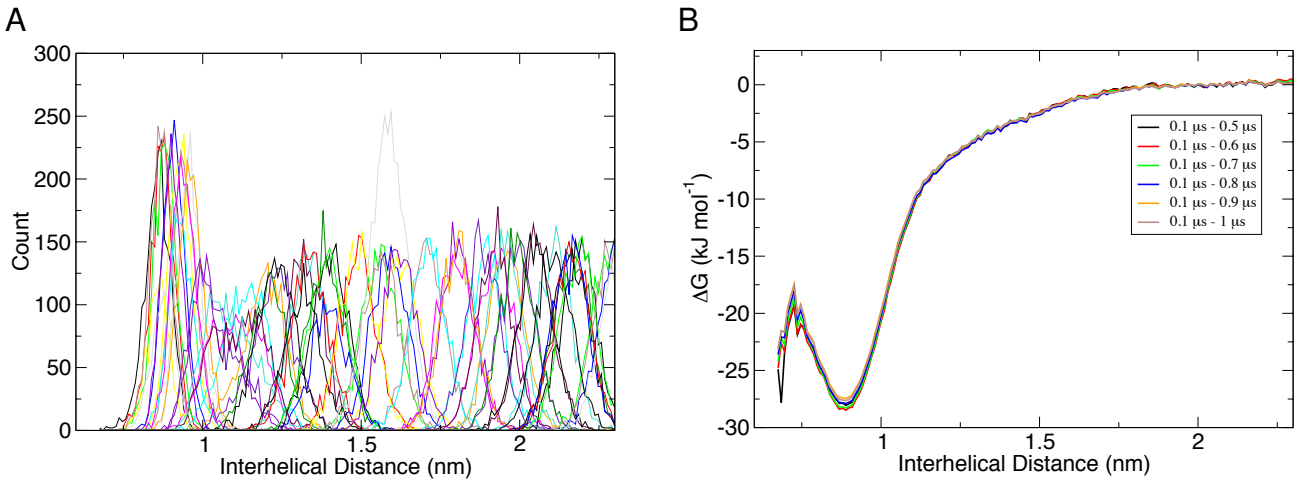

Figure 7.30: (A) AZ2(GG4)$_A$ umbrella sampling histogram coverage of the *inter*helical reaction coordinate. (B) AZ2(GG4)$_A$ converging PMF profiles from 0.1 μs to 1 μs in 0.1 μs increments beginning at 0.5 μs.

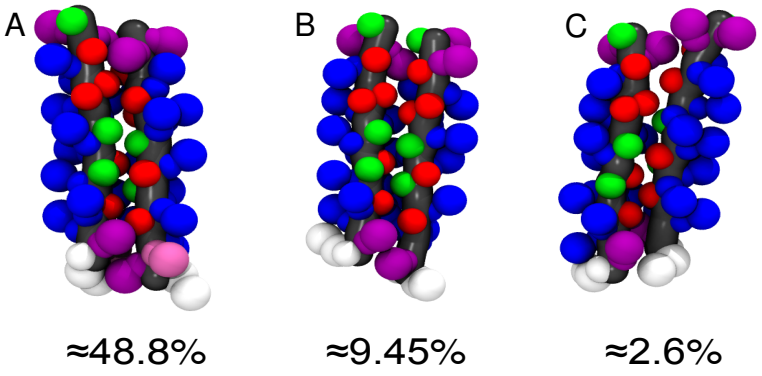

Figure 7.31: Representative of AZ2(GG4)$_A$ dimers from cluster analysis at the free energy minimum found approximately 1 nm along the *inter*helical reaction coordinate. (A) Right helix-helix crossing, (B) zero helix-helix crossing, (C) right-helix-helix crossing. Glycine green, leucine blue, serine purple, tyrosine mauve, alanine red, and arginine white.

equal to ten structures per cluster (Figure 7.32). All eight averaged structures failed to support a right-handed dimer with the GG4 buried at the helix-helix interface as found at the global free energy minimum. All but one of the dimers adopted a left-handed helix-helix crossing orientation, one of which was stabilised by interacting glycine residues. Of the remaining six, two structures presented interlaced leucine residues, whilst the final four presented a variation in alanine with glycine-leucine *inter*helical contacts. The remaining dimer adopted a right handed-helix crossing orientation, with an alanine with glycine-leucine helix-helix interface.

The interaction energy plots relating to the CHI conformational averaged structures are given in Figure 7.33. Although the total interaction energy of structure H yields 149.3 kcal mol^{-1}, the highest interaction energy amongst the eight AZ2(GG4)$_A$ clusters, the N terminal serine and threonine interactions contribute significantly. The six terminal residues, three either ends, are

considered as linker regions of the GALLEX construct and are not meant to contribute towards oligomerisation. Structure G, with a total interaction energy of 125.38 kcal mol^{-1}, returns the second highest level of *inter*helical side chain contribution. Both helical interaction profiles are symmetrical, suggesting a set of homo-interactions, yet interacting side chains do not conform to a repeat of interfacial residues at sites 'a' and 'd'. Rather, the tight packing evident from the observation of structure G is reflected from the addition of the interacting sites 'e' and 'g' in the pattern, $xL^dL^exA^gL^axxL^dL^exA^gL^axxL^dL^e$.

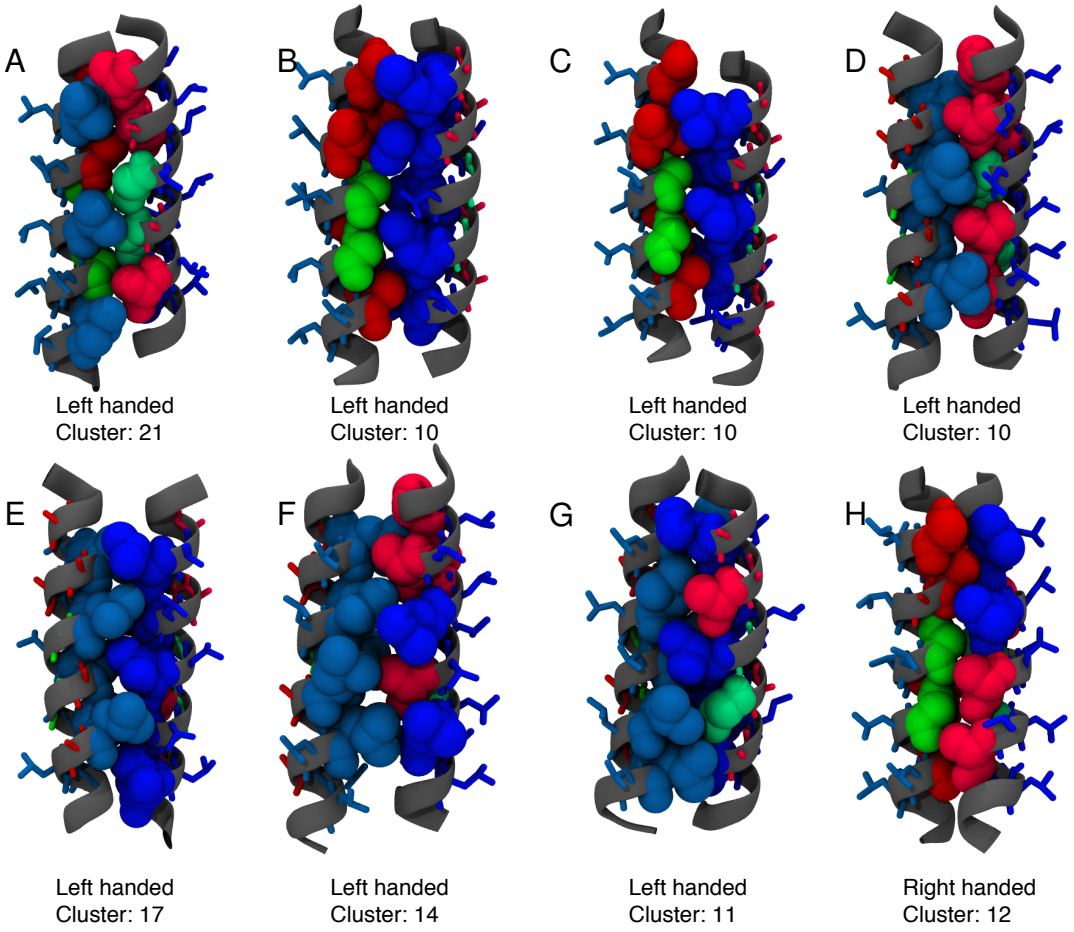

Figure 7.32: Average structures from eight AZ2(GG4)$_A$ clusters. Averaged structures A-G are left-handed dimers. Structure H is a right-handed dimer. For clarity, only residues present on the helix-helix interface are shown in space filling spheres. All other residues are represented in stick-form. Leucine is shown in blue, alanine in red, and glycine in green.

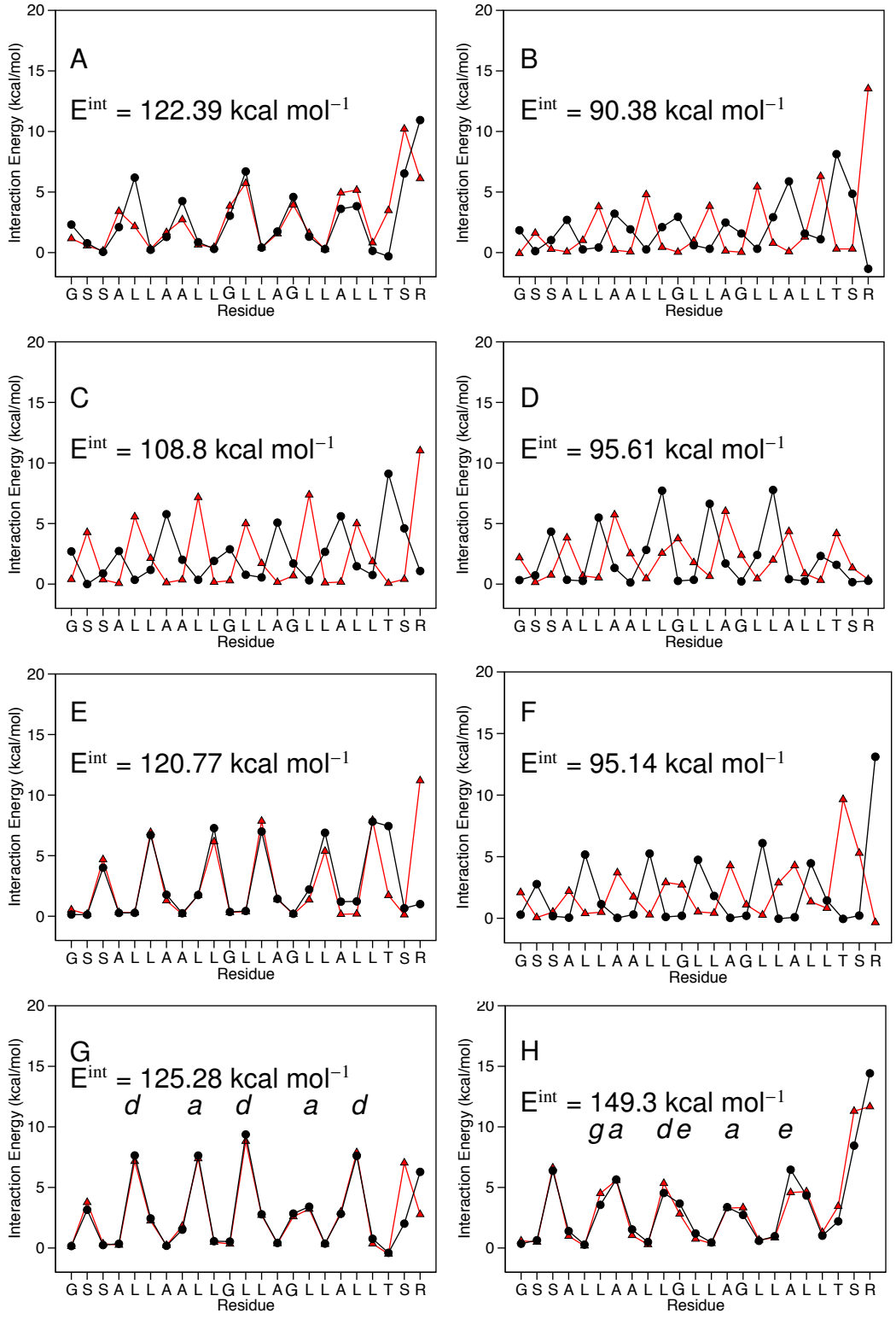

Figure 7.33: Atomistic AZ2(GG4)$_A$ Interaction energies for structures A-H in Figure 7.32. E_{int} indicates the total interaction energy summed over both helices. Each helix is a separate line.

7.4 United atom simulations of AZ2, AZ2L6S and AZ2L10S

The *in vivo* investigation into low complexity sequence oligomerisation yielded a strong association between AZ2L6S and AZ2L10S constructs. The GALLEX methodology alone is unable to explain precisely how the two low complexity sequences have such a high affinity for self-association. Therefore, united atom models of AZ2, AZ2L6S and AZ2L10S were employed to study the helix-helix crossing angle, *inter*helical side chain contacts and the potential for hydrogen bonding between the serine side-chains, and with the serine side chains to backbone. Unlike the previous *in silico* representation of low complexity sequences, a CG model was avoided, as we required a more precise atomic representation to capture potential hydrogen bonds.

7.4.1 Initial dimer geometry

To ensure the dimer configuration was modelled according to the structural analysis from *in vivo* and *in vitro* experimentations in literature (see section 5.1.3 for a detailed explanation) the initial dimer configuration of AZ2L6S and AZ2L10S was selected from a CHI conformation search of structures where a hydrogen bond had potential formed between serine side chain hydroxyl group to the backbone carboxyl oxygen atom (234). In addition, all three dimer sequences must adopt a left handed helix-helix crossing angle to ensure a 'knobs-into-holes' packing, as expected from a heptad repeat dimer (224). An initial crossing angle of 25° was agreed on, given that for the above sequences the crossing angle has yet to be derived from computational or experimental efforts.

An all-atom conformational search using the CHI algorithm was performed using the parameters described in section 7.2.2. Using a 1 Å RMSD cut off, cluster analysis was performed across the search results and an average structure was returned per cluster. All structures found to adopt a right-handed helix-helix crossing angle were disregarded.

7.4.1.1 AZ2 initial dimer geometry

A left-handed dimer with a polyleucine helix-helix interface was selected from the averaged cluster conformations.

7.4.1.2 AZ2L6S initial dimer geometry

Of the averaged structures returned by the CHI clustering algorithm, a single structure satisfied the initial modelling requirements. The inter-digitations of the leucine side chains at the helix-helix interface can be seen in Figure 7.34 (A), along with a clear left-handed helix-helix crossing orientation from the top view (B). The interaction energy (C) between the two helices reflects the symmetrical nature of the interacting side chains. For example, in the case of a homodimer, interaction between identical sets of interfacial side chains would yield two identical potential energy profiles. The hydroxyl group on each serine side chain was within 4 Å of a backbone carboxyl oxygen atom as identified using the VMD visualisation package.

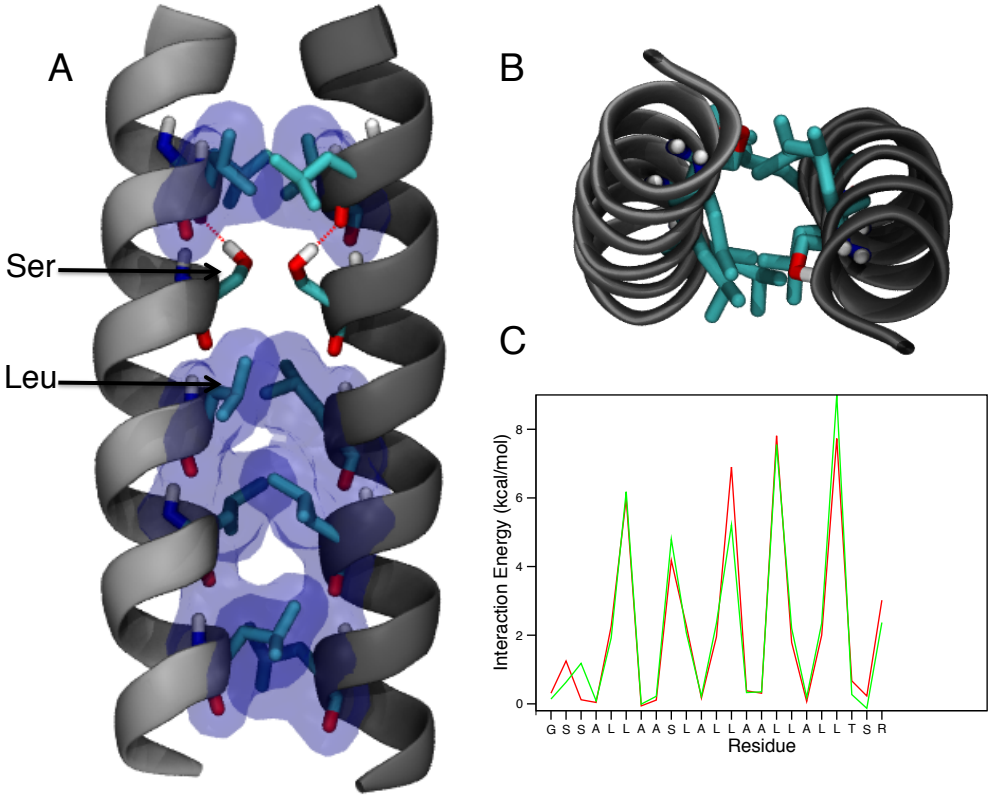

Figure 7.34: AZ2L6S. (A) Packed side chains along the helix-helix interface. (B) Top down view of the left-handed crossing angle. (C) Interaction energy of each helix.

7.4.1.3 AZ2L10S initial dimer geometry

Unlike AZ2L6S, the most suitable left-handed averaged structure of AZ2L10S (Figure 7.35 (A)) did not adopt a symmetrical helix-helix interface, rather the arrangement of interacting side-chains were off-set by a residue. This is reflected by the interaction energy presented in Figure 7.35 (C).

The hydroxyl group on each serine side chain was within 4 Å of a backbone carboxyl oxygen atom as identified using the VMD visualisation package

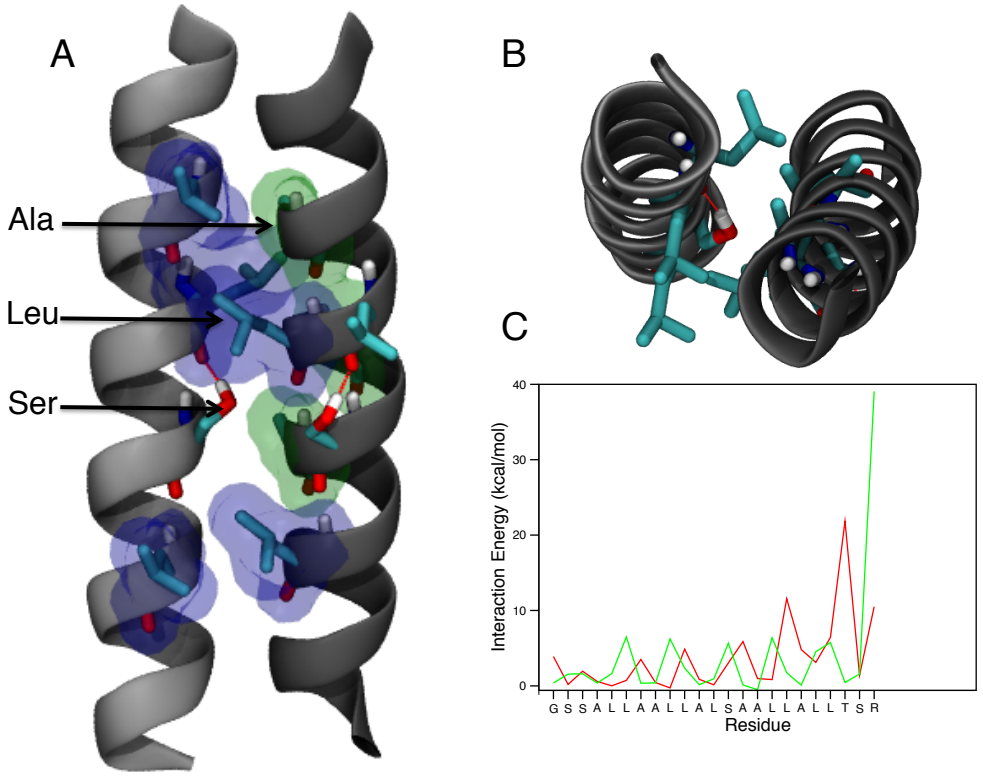

Figure 7.35: AZ2L10S. (A) Packed side chains along the helix-helix interface. (B) Top down view of the left-handed crossing angle. (C) Interaction energy of each helix.

7.4.2 Bilayer area per lipid

Dimers AZ2, AZ2L6S, and AZ2L10S in POPC bilayers were equilibrated according to section 7.2.3.6. To ensure that each bilayer had equilibrated the bilayer area per lipid was recorded during the equilibration stages in addition to a 45 ns production simulation. The calculations for AZ2, AZ2L6S, and AZ2L10S are presented in Figure 7.36, 7.37, and 7.38, respectively. As a result of iterative compression and energy minimisation steps, the initial area-per-lipid for all three systems came within approximately 2 $Å^2$ of the reported experimental area-per-lipid. Systems were left to equilibrate over 20 ns. Position restraints were gradually lifted from the protein backbone. The area-per-lipid for AZ2 was shown to drop towards its average, where the bilayer area-per-lipid AZ2L6S and AZ2L10S began converge within 20 ns. The final calculation (note, the initial 5 ns on unrestrained simulation was given as additional time for bilayer equilibration) was taken as

an average over 50 ns along with the standard deviation, reported as a blue solid line and blue dashed line, respectively. Each figure is presented with the experimental area-per-lipid of POPC (67.4 $Å^2$ ± 1.3 $Å^2$) according to small-angle neutron and X-ray scattering (253). The three average area-per-lipid calculations came within 4 $Å^2$ of the reported experimental values of pure POPC. The difference should be considered a software limitation due to an approximation of the area-per-lipid when a portion of the bilayer has been taken up by the presence of a protein. Given that the area-per-lipid calculations during the production simulation did not drastically deviate from the average value, it was within reason to assume that structural analysis could be performed on each dimer over the last 45 ns of unrestrained simulation time.

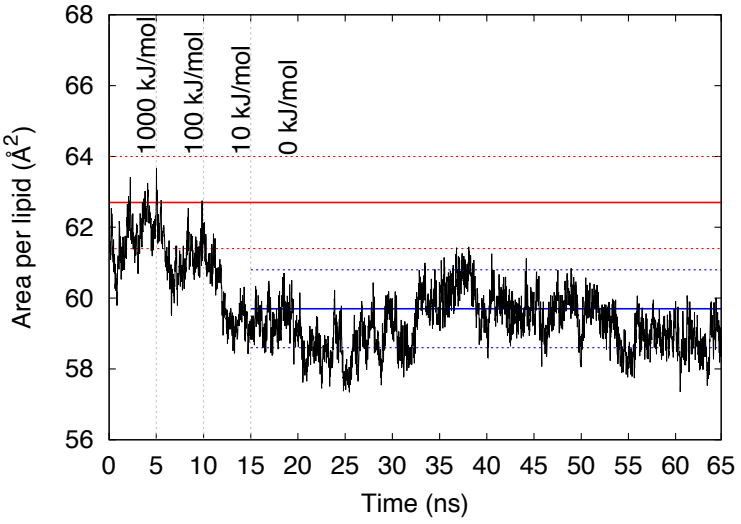

Figure 7.36: Area-per-lipid of the lipid environment surrounding the AZ2 dimer. Experimental average with standard deviation is depicted in red. The average area-per-lipid calculated across the production run is depicted in blue.

7.4.3 *inter*helical distance calculations and serine hydrogen bonding

As seen in Figure 7.39 (A), AZ2 had adopted a predominately leucine packed helix-helix interface with leucine pairs: L_7-L_7; L_{14}-L_{14}; and L_{17}-L_{17}, and leucine-alanine pairs: A_4-L_3; L_4-A_3; L_7-A_8; A_8-L_7; L_{10}-A_{11}; and A_{11}-L_{10}. Represented by the final time frame in Figure 7.40 (A), AZ2L6S adopts a left handed helix-helix crossing orientation with leucine side chains packing at the helix-helix interface, similar to the packing arrangement of a leucine zipper. The orientations of the helices have clearly remained stable compared with their initial configuration, with the exception of the position of the serine side chain adjusting to a distance less than 4 Å , within the potential cut-off for hydrogen bonding (as seen in Figure 7.40 (A) by a black line between the oxygen

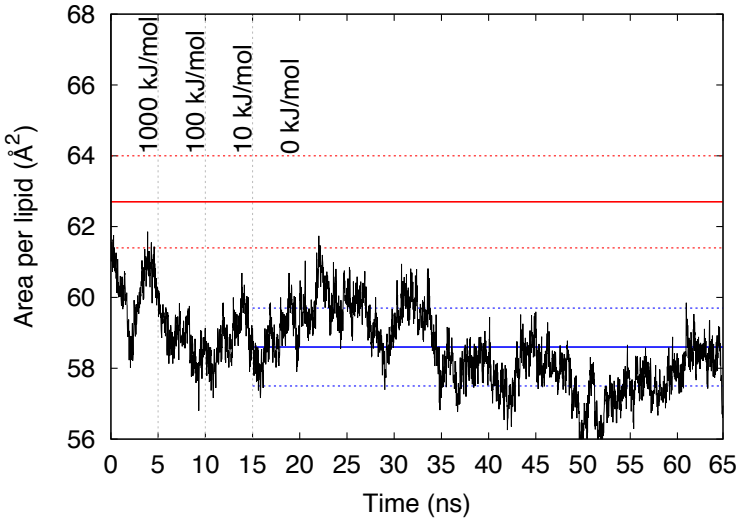

Figure 7.37: Area-per-lipid of the lipid environment surrounding the AZ2L6S dimer. Experimental average with standard deviation is depicted in red. The average area-per-lipid calculated across the production run is depicted in blue.

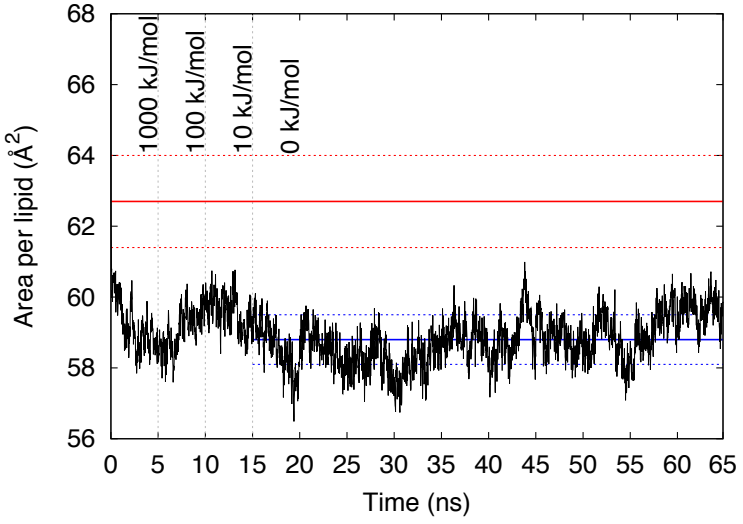

Figure 7.38: Area-per-lipid of the lipid environment surrounding the AZ2L10S dimer. Experimental average with standard deviation is depicted in red. The average area-per-lipid calculated across the production run is depicted in blue.

acceptor and hydrogen donor). The *inter*helical helix-helix side chain pairs: L_2-L_3; L_3-L_2; S_6-S_6; L_9-L_{10}; L_{10}-L_9; and L_{13}-L_{14}; L_{14}-L_{13}; and L_{17}-L_{17}, are on average well within interaction distance (see Figure 7.40 (B)). Unlike AZ2L6S, the AZ2L10S dimer presents a more asymmetrical set of *inter*helical interaction pairs, with alanine being well within potential interaction distance of leucine. The pair of helices retained the left-handed helix-helix crossing angle (see Figure 7.41 (A)) and the interaction pairs: L_3-A_4; L_3-L_3; L_6-A_8; L_7-L_7; S_{10}-S_{10}; L_{13}-A_{15}; L_{14}-L_{14}; and L_{17}-L_{17}, are on average, well within potential interaction distance (see Figure 7.41 (B)). During the course of the production run, a serine-serine *inter*helical hydrogen bond is favoured over the initial

serine-to-peptide backbone hydrogen bond initially presented in Figure 7.35 (A).

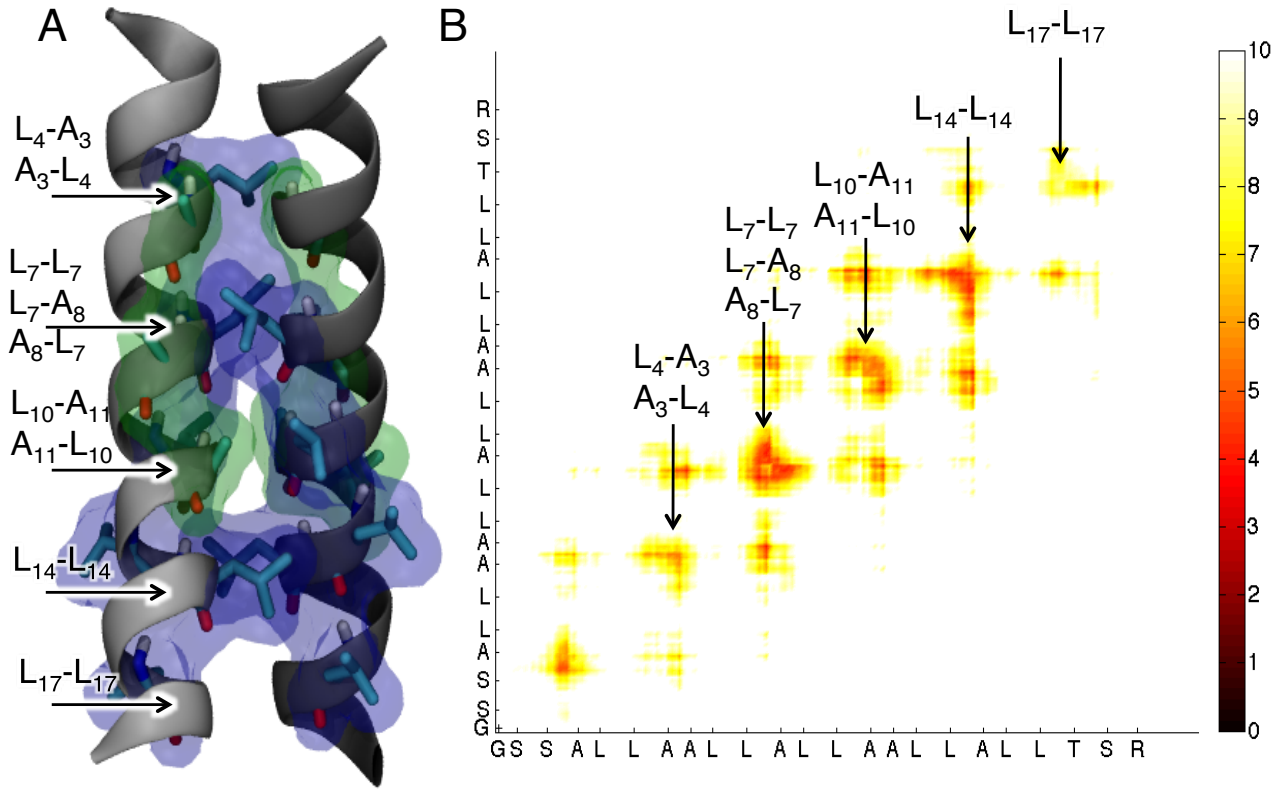

Figure 7.39: (A) The final time frame from the production run of AZ2. *inter*helical side chains within 10 Å are depicted in semi-transparent spheres. Leucine in blue, and alanine in green. (B) The averaged *inter*helical residue atom-atom distance. Only atoms within 10 Åare displayed. Interfacial residues are labeled.

7.4.4 Helix-helix crossing angle

The helix-helix crossing angles of AZ2, AZ2L6S and AZ2L10S were calculated, according to the method presented in Appendix A, over the last 45 ns of production run with the initial 20 ns taken up by system equilibration. The error was reported as the standard deviation. The average crossing angle as presented in Figure 7.42 are 41.54 ± 2.51°, 24.07 ± 2.56°, 25.77 ± 1.83° for AZ2, AZ2L6S, and AZ2L10S, respectively. Each dimer had begun with an identical crossing angle prior to equilibration.

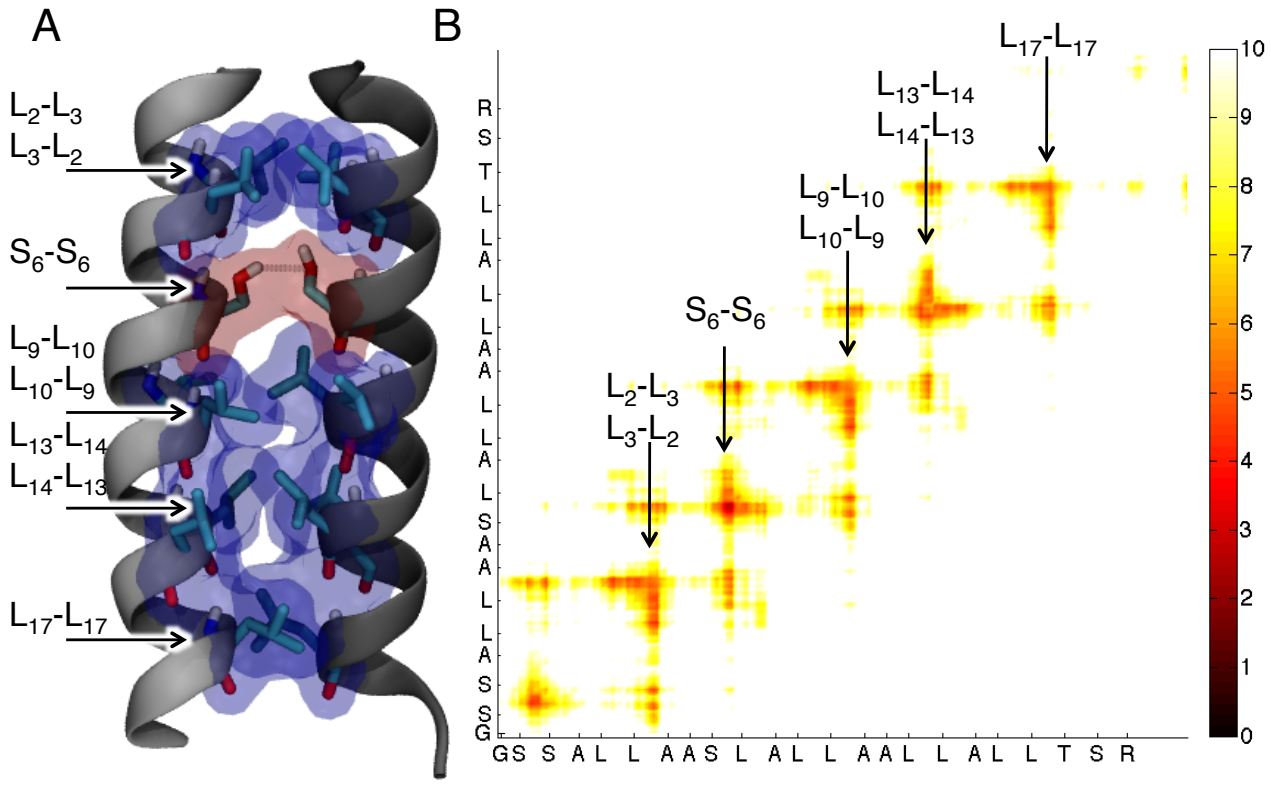

Figure 7.40: (A) The final time frame from the production run of AZ2L6S. *inter*helical side chains within 10 Å are depicted in semi-transparent spheres. Leucine in blue, and alanine in green. (B) The averaged *inter*helical residue atom-atom distance. Only atoms within 10 Å are displayed. Interfacial residues are labeled.

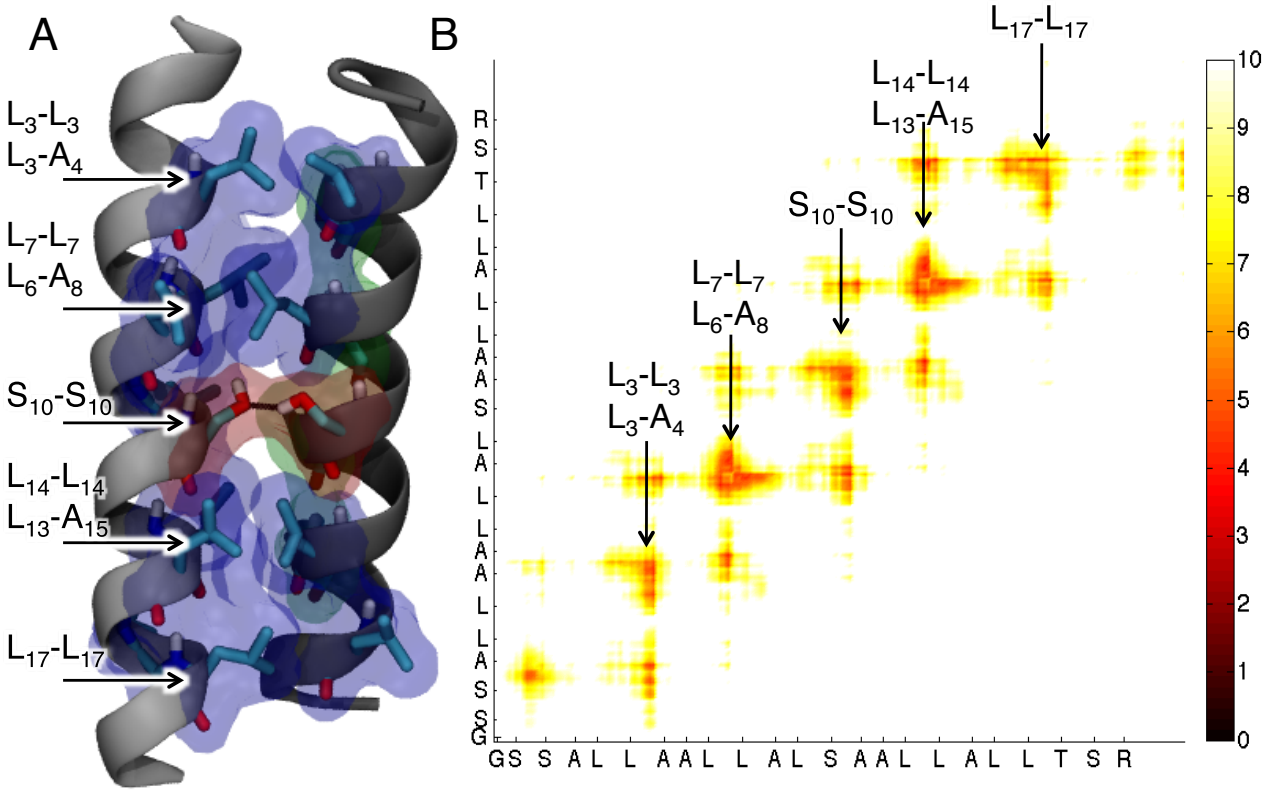

Figure 7.41: (A) The final time frame from the production run of AZ2L10S. *inter*helical side chains within 10 Å are depicted in semi-transparent spheres. Leucine in blue, and alanine in green. (B) The averaged *inter*helical residue atom-atom distance. Only atoms within 10 Åare displayed. Interfacial residues are labeled.

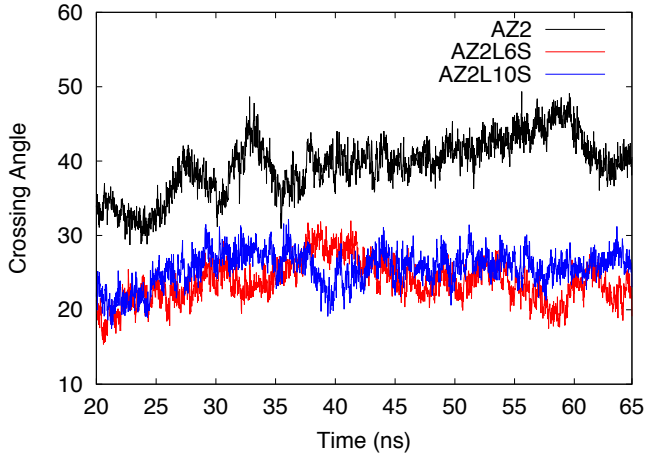

Figure 7.42: The average crossing angle with standard deviation calculated over a 45 ns production run. AZ2: 41.54° ± 2.51°, AZ2L6S: 24.07° ± 2.56°, and AZ2L10S: 25.77° ± 1.83°.

7.5 Discussion

This chapter reports free energy calculations of self-association, helix-helix crossing orientations and the packing motifs of interfacial side chains of the low complexity scaffolds presented earlier in chapter 5.

The global free energy of self-association was determined for the low complexity sequences: L17; L17(GG4); L17L9Q; AZ2; AZ2$(GG4)_L$ and AZ2$(GG4)_A$. Potential of mean force calculations were obtained using umbrella sampling across an *inter*helical reaction coordinate between CG helices. The trajectory of the umbrella window of which the *inter*helical distance matches the global free energy location on the reaction coordinate, was subjected to cluster analysis. A representative dimer from each cluster was reported. The helical composition for all sequences have yet to reported in literature. The TM peptides were assumed to form α-helical structure due to the propensity for leucine and alanine to form helical structures in a non-polar environment (256). Therefore, the inability of the CG model to change the helical content was of no concern at this early stage of investigation. Our early work on free energy calculations of TM domain association (chapter 4) made it clear that due to long autocorrelation times of membrane systems (198), each umbrella window required at least 40 ns of simulation time, of which equilibration accounted for at least 10 ns to 20 ns. CG calculations appeared to be a good compromise given that we had the *in vivo* experimental approximate free energy values to compare against.

The CG simulations were accompanied by an atomistic conformational search using CHI, a tool for exploring potential TM helical packing interfaces. Given the reduced degrees of freedom in the CG model, there was some concern over whether a CG representation of leucine could replicate the 'knobs-into-holes' packing of leucine residues in a heptad repeat sequence (54, 60). The cluster analysis of atomistic conformations return interaction plots, which conformed, to a heptad repeat pattern. Yet there is no penalty for orientating polar residues away from the packing interface, and without environmental pressure such as hydrophobic mismatch, a conformational search can only yield a side chain packing motif as a result of side-chain interaction energies only. Nonetheless, CHI structures can be used to elicit contributing side chain interactions after a successful comparison has been made with the most prevalent CG structure at the global free energy minimum.

Due to the large difference in GALLEX signal between AZ2L6S and AZ2L10S (Figure 5.9),

additional CHI simulations were performed and select averaged structures were converted into a united atom force field model and subjected to MD simulations. This level of granularity enables the detection of potential hydrogen bonds.

The changes in free energy between low complexity scaffolds and the low complexity sequences after including a motif have been recorded as a flow chart in Figure 7.43. After the inclusion of a helix-helix interaction motif, a decrease in free energy, subsequently resulting in a stronger dimer, is depicted as a green arrow. Conversely, a red arrow depicts an increase in free energy. A black arrow indicates no changes in free energy, or at least $\Delta\Delta G$ is within error. The helix crossing orientation is also included.

7.5.1 L17, L17(GG4) and L17L9Q

The global free energy minimum of L17 was approximately -16.75 ± 0.4 kJ mol^{-1} and shown to predominately adopt a left helix-helix crossing orientation. The trajectory of each umbrella window presents a fully inserted TM peptide, confirming the propensity to spontaneously insert into a bilayer supporting maltose minimal media assay results in section 5.3.1.3 and the hydroxide washes in section 5.3.1.2. Although the *inter*helical distance umbrella restraint between TM peptides in the CG models would prevent deviation from a specified *inter*helical distance, the potential would not prevent either peptide from moving out of the bilayer if its presence was unfavourable. All identified clusters of a CHI atomistic conformational search returned left-handed dimers (Figure 7.4). Further analysis of these structures revealed that the most favoured, according to the total side chain interaction energies (Figure 7.5), clearly adopt a heptad repeat pattern of interacting sites 'a' and 'd'. This is typical of a leucine-zipper like motifs (54). Although it would be very presumptuous to compare free energy of association values between the experimental GALLEX assay and the CG PMF profile, both techniques do indicate that L17 will self-associate, marginally.

After replacing L_7 and L_{11} with glycines, the ΔG of self-association dropped to -19.16 ±0.4 kJ mol^{-1} (represented by a green arrow on the thermodynamic pathway in Figure 7.43). The trajectory of each umbrella window presents a fully inserted TM peptide, supporting maltose minimal media assay results in section 5.3.1.3 and sodium hydroxide washes in section 5.3.1.2. Cluster analysis at the global free energy minimum suggested a propensity for a left helix-helix crossing orientation with the GG4 motif buried at the helix-helix interface (Figure 7.8). The location of the global

free energy minimum differed by -0.1 nm compared with L17, as a consequence of closer packing of the helices due to the short side chain of glycine. This can be seen from the identified CG clusters A-C. Structures D and E present a leucine packed helix-helix interface, however, this should not diminish that fact that the remainder of the trajectory (not grouped into clusters of a significant size) fluctuated around a region of phase space where the GG4 was packed at the interface. Further more, GG4 mediated packing of native TM peptides suggest that the packed glycines present the shortest *inter*helical distance between back bones (32, 30). The models presented by CHI, do not immediately match with the CG GG4 mediated dimers. Rather, they present left-handed dimers conforming to a heptad repeat orientation. CHI does not take into account the borderline hydrophobicity of glycine as indicated by the theoretical free energy values presented in section 5.3.1.1, which predicted that inserting glycine into the bilayer was unfavourable. In addition, the CG model would take into account a drop in possible side chain conformations if leucine were to pack at the helix-helix interface. Therefore, we suggest from the comparison between experimental results and the theoretical and computation models, that L17(GG4) has a propensity to form a right-handed GG4 mediated dimer, with a possible transition to a left-handed leucine mediated heptad repeat dimer.

It is interesting to note that although both experimental and CG free energy values indicate that L17(GG4) can self-associate, they do not agree with the difference in free energy between L17(GG4) and its predecessor, L17. Whereas the GALLEX assay indicates that introducing a GG4 motif into a polyleucine low complexity scaffold destabilises oligomerisation, the CG model suggests that it stabilises dimerisation.

After replacing the L_9 of L17 with a glutamine the ΔG increased to -10.5 ±1 kJ mol^{-1}, suggesting a significant drop in propensity for self-association (represented by a red arrow on the thermodynamic pathway in Figure 7.43). However, the global free energy minimum still reports an energetically favourable dimer. The trajectory of each umbrella window presents a fully inserted TM peptide, supporting maltose minimal media assay results in section 5.3.1.3 and sodium hydroxide washes in section 5.3.1.2. The PMF profile presents two free energy minimum, of approximately the same ΔG. Four clusters of right handed dimers were identified from the trajectory best matching the free energy minimum at approximately 0.95 nm along the reaction coordinate.

One has to question the energetic contributions, whether favourable or not so, of the position of

the glutamine residue. Approximately 0.95 nm along the reaction coordinate, a leucine packed helix-helix interface is favoured (Figure 7.13). Considering the difference between free energy minima of L17 and L17L9Q, and the leucine packed L17 helix-helix interface, the difference in free energy can be attributed to the position of the glutamine residue into the bilayer. Of the eleven CHI conformational search structures (Figure 7.15), only structure G presented a potential interaction between glutamine side chain. The remaining structures presented a leucine packed helix-helix interface with a left-handed helix crossing orientation. Interestingly, the only CHI structures adopting a right handed dimer (J and K) yield the least energetic contributions from interacting side chains. Given that the CG free energy minimum presents in abundance of right-handed dimeric structures, these two models do not compare well.

At the second free energy minimum at 0.5 nm on the reaction coordinate, cluster analysis presented dimers with the glutamine side chain buried within the helix-helix interface (Figure 7.14), whilst adopting a mix of left, right and zero helix-helix crossing orientations. The amount of variation in representative CG structures compares well with the amount of variation in the CHI models (Figure 7.15).

The depth and value of both free energy minima suggest a complicated free energy landscape, unseen in either L17 or L17(GG4). There is a significant energy penalty for exposing glutamine to the hydrophobic lipid chains (66), therefore burying the polar side chain into the helix-helix interface would mediate one state of self-association. On the other hand, there are ample contributions to helix-helix interactions from leucine side chains which not only mediates self-association but also compensates for glutamine's energetic penalty.

Both experimental and CG free energy calculations suggest that L17L9Q self-associates, which compares well with helix-association of L_7QL_{15} peptides in detergent micelles analysed using SDS-PAGE (70). The results in our study differ, however, when the free energy of L17L9Q is compared with its predecessor L17. GALLEX experimental data in section 5.3.3 suggests that introducing a glutamine residue improves the stability of a polyleucine low complexity scaffold, whilst the CG model suggest otherwise. Prior investigation by Zhou *et al.,* (70) concluded that a polyleucine peptide was predominately, although not completely, monomeric whilst a L_7QL_{15} peptide was marginally more dimeric. It is likely that the difference in self-association between the two techniques relative to a polyleucine control is the result of the mimetic-environment and the

CG model itself, losing the potential of forming hydrogen bonds between glutamine side chain.

7.5.2 AZ2, AZ2(GG4)$_{L}$ and AZ2(GG4)$_{A}$

The free energy of self-association as a function of *inter*helical distance between two CG AZ2 low complexity scaffolds was -24.75 ± 0.8 kJ mol^{-1}. AZ2 forms a significantly more stable dimer than the CG polyleucine low complexity scaffold. The CG free energy value, relative to L17, compares well with the GALLEX assay experimental results in section 5.3.3, in that AZ2 *in vivo* has a stronger propensity for self-association than L17. With a difference in free energy by as much as 8 ±1.2 kJ mol^{-1}, the free energy calculations of the CG models compare well to what is almost a one third difference reported in POSYYCCAT signal between self-association of polyleucine and AZ2 (104). The trajectory of each umbrella window presents a fully inserted TM peptide, supporting maltose minimal media assay results in section 5.3.1.3.

Cluster analysis performed at the global free energy minimum indicates that approximately 98% of the recorded trajectory adopted a left-handed dimer with alanine residues packing at the helix-helix interface (Figure 7.20). To the author's knowledge, there is no structural data with regards to the precise packing motif of an AZ2 dimer. According to the theoretical free energy of TM peptides spontaneously inserting into the bilayer (section 5.3.1.1), alanine is significantly unfavourable, whilst leucine being a very hydrophobic residues is considered as very favourable. This would suggest that AZ2 dimers would favour an alanine packed helix-helix interface, given the borderline hydrophobicity of alanine (220). In addition, the constrained side chains of alanine, may limit the entropic cost of TM domain association (28).

A conformational search of an atomistic AZ2 dimer using the CHI algorithm returned eight clusters (Figure 7.21). Seven of the eight clusters returned a left-handed dimer. Although there are no lipid or solvent contributions to side chain packing in a CHI model, the search yielded two clusters of left-handed dimers with an alanine side chain packed helix-helix interface. According to the total interaction energy (Figure 7.22), neither structure is capable of maximising on interaction energy from alanine side chain interactions alone. This would suggest that the additional contributions required to maintain an alanine packed helix-helix interface at the CG free energy minimum are potentially lipid mediated.

The free energy of AZ2(GG4)$_{L}$ self-association was -24±1 kJ mol^{-1}, statistically equivalent to

AZ2. Therefore, as indicated in the thermodynamic pathway (Figure 7.43) there is no change to the propensity for self-association after the inclusion of a GG4 motif amongst leucine residues. The trajectory of each umbrella window presents a fully inserted TM peptide, supporting maltose minimal media assay results in section 5.3.1.3. Although the GALLEX free energy assay results presented in section 5.3.3 suggests a marginal increase in self-association, the ever so slight difference in free energy value relative to AZ2 between techniques does not imply that the CG results of AZ2$(GG4)_L$ are of little interest. On the contrary, the CG AZ2$(GG4)_L$ global free energy minimum presents a low complexity sequence with a propensity for self-association similar to the AZ2 scaffold, but results in a totally different dimeric orientation.

Cluster analysis at the global free energy minimum of AZ2$(GG4)_L$ returned a startling contrast in dimeric conformations to the AZ2 clusters (Figure 7.25). Unlike the left-handed AZ2 dimer, where alanine side chains were found to pack at the helix-helix interface, a GG4 motif associating with neighbouring leucines encouraged a right-handed dimer with leucine side chains in addition to the GG4 motif packing at the interface. The right handed helix-helix crossing orientation, as is the case in GpA, is indicative of a GG4 mediated dimer (30). AZ2 and AZ2$(GG4)_L$ have approximately similar global free energy minimum, and yet the orientation of the flexible leucine residues switch. No longer is the constrained side chain of alanine packing to compensate for a potential entropic cost of association. It would be within reason to suggest that the GG4 mediated dimerisation leads to a tighter arrangement of interacting leucine side chains, however, a CHI conformational search was unable to yield an appropriate structure and subsequent interaction profile.

The self-association global free energy minimum of AZ2$(GG4)_A$ was -27.4±0.4 kJ mol^{-1}, 2 ± 1.2 kJ mol^{-1} more favourable than AZ2 (represented by a green arrow on the thermodynamic pathway in Figure 7.43). The trajectory of each umbrella window presents a fully inserted TM peptide, supporting maltose minimal media assay results in section 5.3.1.3.

Over 60% of the trajectory at the global free energy was found to adopt a right handed dimer, stabilised by alanine and glycine side chain interactions (Figure 7.31). As with AZ2$(GG4)_L$, the AZ2$(GG4)_A$ is seen to switch crossing orientation with respect to the AZ2 low complexity scaffold. Only, there is now a distinct increase in dimer stability. This compares well to the free energy calculations from GALLEX assays presented in section 5.3.3. Interestingly, a CHI conformational search was able to return a right handed dimer (structure H of Figure 7.32), of which the contributing

side chain interactions at the helix-helix interface were predominately leucine and glycine. Their energetic profile was not only highest in energy of the isolated clusters, but also conformed to a heptad repeat sequence of interacting ‘a’, ‘d’, ‘e’ and ‘g’, side chains.

7.5.3 United atom simulations of AZ2, AZ2L6S and AZ2L10S

In contrast to earlier works by Gray and Matthews (234) and Ballesteros *et al.,* (235), the GALLEX signals from oligomerisation of TM domains AZ2L6S and AZ2L10S, and as discussed in section 5.4.2, suggest that a single serine residue can prompt oligomerisation. Therefore, in order to understand the role serine plays in TM domain dimerisation, united atom models of AZ2, AZ2L6S and AZ2L10S were simulated for 50 ns in a solvated POPC bilayer. The crossing angle, *inter*helical side chain distances and potential for hydrogen bonding was analysed. Initial structures of AZ2L6S and AZ2L10S were selected from a CHI conformational search of left-handed structures due to the assumed hydrogen bonding potential between the hydroxyl group on the serine side chain and the carboxyl group on the same helix (234). Although a 50 ns production simulation is typically far too short to capture unfolding-folding events of a helical structure in a bilayer, the presumed *inter*helical hydrogen bonding of serine was shown to shift into the helix-helix interface and came within the potential distance required to establish an *inter*helical hydrogen bond between the opposite serine side chain.

The initial dimer structures were selected from a CHI conformational search which had been constrained to a left-handed helix-helix crossing angle of 20°. During the equilibration phase the crossing angle of AZ2 was seen to increase by approximately 12°, from which point the crossing angle continued to increase over the production run until converging to an average of 41° ± 2.51°. On the other hand, both AZ2L6S and AZ2L10S were shown to increase by approximately 5° during the early stages of the production simulation until converging to an average of 24° ± 2.56° and 25° ± 1.83°, respectively. The enthalpic contributions from the formation of a hydrogen bond between serine side chains, could encourage a reduction in the helix-helix crossing angle relative to the crossing angle of the AZ2 control, encouraging a close packing along the length of the helix-helix interface. Such a mechanism may explain the increase in oligomerisation when a serine is substituted into AZ2. If a larger crossing angle was to form (similar to the crossing angle of AZ2), the potential for *intre*helical side chain formation between serine residues is reduced due to

the increased distance between hydrogen donor and acceptor groups.

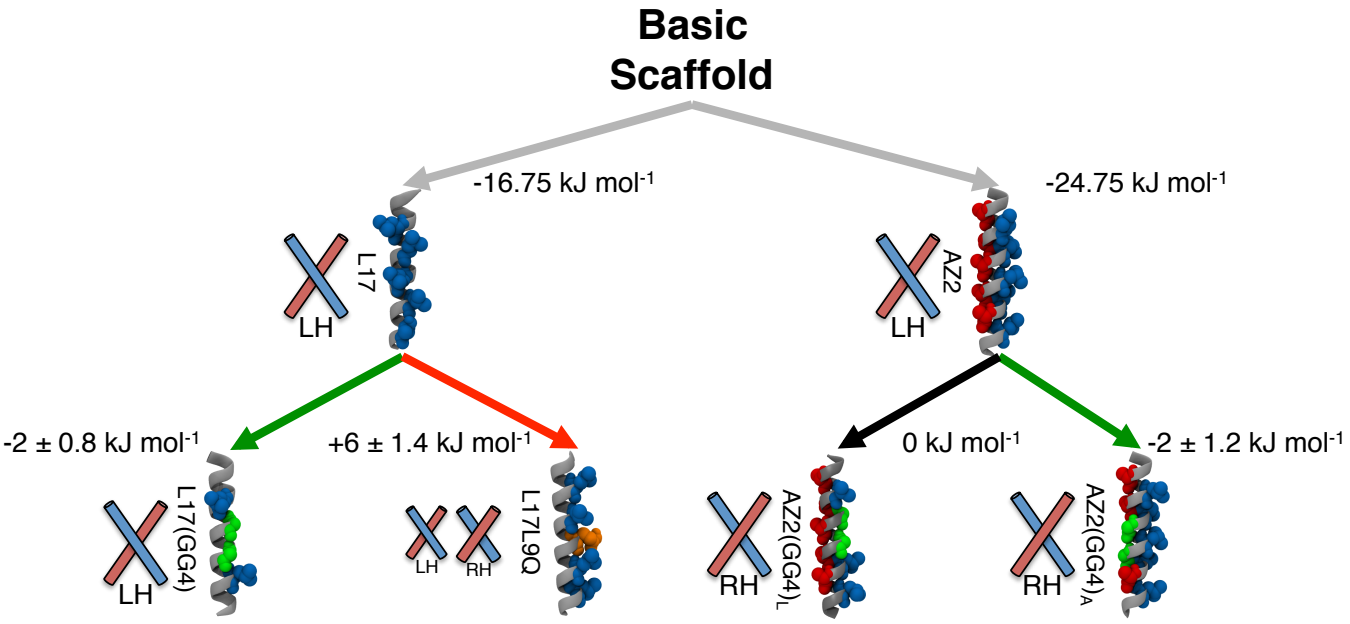

Figure 7.43: Thermodynamic pathway of CG basic scaffold sequences. Change in free energy from respective low complexity scaffolds are reported. In the case of AZ2 and L16, the change is relative to 0 kJ mol^{-1}. Beside each helical representation is an illustration of the predominant helix-helix crossing orientation located at free energy minimum and supported by side chain interaction profiles from CHI conformational clustering analysis. Sequences which adopt a mix of left, right and zero crossing angles are presented by a left and right illustration. Residues which make up the low complexity sequences and scaffolds are represented in space filling spheres with leucine in blue, alanine in red, glutamine in orange and glycine in green.

7.6 Conclusion

Low sequence complexity TM domains featuring key helix-helix interaction motifs were modelled using CG molecular dynamics. The AZ2(GG4)$_A$ sequence was found to be the most stable dimer. In addition, a GG4 motif was capable of switching the left-handed AZ2 dimer into a right-handed dimer, whether associating with the leucine face or alanine face, sequences AZ2(GG4)$_L$ and AZ2(GG4)$_A$ respectively. With an almost equal propensity for dimerisation, the switch between crossing angle orientations may serve as a basis for a *de novo* TM switch mechanism. Further more, it may introduce new insight behind oncogenesis of receptors of which the mechanism of activation is thought to be controlled via conformational change (183, 179). Finally, there is general agreement between the free energy rank order and the GALLEX free energy calculation. Although there is some discrepancy behind the true contribution to self-association from glutamine interactions, which given the difficulty of comparing biological membranes with model membranes is likely to be a lipid-mediated discrepancy.

Given the propensity for oligomerisation associated with the sequences AZ2L6S and AZ2L10S, a united atom MD simulation was performed on their TM domain. From initial starting conformation with serine's potential for hydrogen bonding solely with the backbone carboxyl group it was interesting to report that after a series of equilibration steps and a production run the hydrogen bond had potential switched to a serine-serine configuration. Further to this chapter, it would be insightful to see how a potential hydrogen bond between the two helices of AZ2L6S and AZ2L10S could contribute to the energetics behind self-association.

Chapter 8

Conclusion and Future Work

8.1 Conclusions

In chapter 1, the widely known characteristics of TM proteins and the simplified two-stage model of membrane protein insertion and folding were introduced. This was followed by a review of interaction motif mediated dimerisation, notably from heptad repeat sequences, GG4 motifs and polar residues. An in depth review on low complexity sequences found in the literature followed. It is worth noting the limited number of MD studies with respect to the number of experimental studies. MD is not an ideal platform for studying a large number of sequences, unless the computational cost can be reduced by either using an implicit bilayer or a CG model (or significant computational resources are readily available). Rather, a library of plasmids can be used to rapidly screen TM sequences using *in vivo* TM interaction reporter assays before investing time into computational studies. The simulation theory and experimental techniques were outlined in chapter 2 and chapter 3, respectively.

The results from 'top-down' MD investigations of the Neu receptor tyrosine kinase were presented in chapter 4. A 'bottom-up' investigation of TM domain low complexity sequence designs using the *in vivo* GALLEX assay (26) were reported in chapter 5. This was followed in chapter 6, by an investigation on the variation between PMF profiles of CG and united atom simulations of TM protein models. Finally, chapter 7 presents a 'bottom-up' investigation of the aforementioned low complexity sequences in order to elicit dimeric crossing orientation and side chain packing motifs

at the free energy minimum. In what follows, we present a very brief summary of the key results as expressed in further detail in the discussion section of the chapters 4, 5 and 7.

In chapter 4, a 'top-down' investigation of the Neu TM protein domain revealed that the *inter*helical approach is a lipid mediated process, driven by the hydrophobic mismatch between the Neu TM protein domain length, and bilayer thickness. A similar study into the bilayer thickness as a function of *inter*helical distance between ErbB-2 was reported by Prasanna *et al.,* (216). Their model, however, was CG which fails to model variation in secondary structure, hydrogen bonding and as seen in chapter 6, a full PMF profile as a consequence to long range lipid mediated interactions. Out results show that both helical orientations of Neu and Neu* have a similar propensity for dimerisation, although self-association of Neu* may prematurely halt beyond short-range protein-protein interactions due to the propensity for the glutamic acid substitution to interact with the lipid head group region. Finally, both Neu and Neu* dimers were stabilised by side chain interactions. The *inter*helical distance between V_{664}E substitutions during association of A_{661}xxxG_{665} interfaces, suggests the formation of hydrogen bonding potentially supporting the model proposed by Gullick *et al* (73). The exposed polar side chain of V_{664}E during association of I_{659}xxxV_{663} interfaces was found to decrease the cost of hydrophobic mismatch by forming a hydrogen bond with the peptide backbone and orientating towards the lipid-water interface.

The 10 μM induced GALLEX assays of low complexity sequences presented in chapter 5 yielded an order of propensity for oligomerisation in a biological context: AZ2L10S > $AZ2(GG4)_A$ > AZ2L6S ≈ $AZ2(GG4)_L$ > AZ2 > L17L9Q > L17 ≈ L17(GG4). These results were complemented using IPTG titration measurements, which revealed the apparent ΔG of association. From the values obtained, a thermodynamic pathway was constructed revealing an order of the most favoured sequences: $AZ2(GG4)_A$ > $AZ2(GG4)_L$ > AZ2 > L17L9Q > L17 >≈ L17(GG4). The experimental free energy results were complemented by CG molecular dynamics and atomistic conformational searches using CHI (191). Global free energy minima from PMF calculations yielding the ΔG order of: $AZ2(GG4)_A$ > AZ2 ≈ $AZ2(GG4)_L$ > L17(GG4) > L17 > L17L9Q. It is clear that dimer variants of AZ2 were as stable if not more stable than AZ2 itself, and significantly more stable than L17 (which formed a very weak dimer regardless to whether a single glutamine or GG4 motif was included). Yet, the order of experimental and computational free energy calculations of the L17 sequences did not match precisely. This is most likely due to differences between the composition

of a POPC model membrane to a bacterial inner membrane, which is discussed in sight of the limitations of this study below. One of the most interesting results is the switch in dimer crossing orientation after a GG4 motif has been inserted into the leucine or alanine face of AZ2. When glycines are placed approximately halfway along the TM domain, the left handed AZ2 dimer adopts a right handed crossing orientation.

Finally, results from GALLEX signals and observations from united atom simulations suggest that dimerisation of AZ2 can be stabilised further by a potential *inter*helical interactions of serine. This is in stark contract with the initial hypothesis by Gray and Matthews (234) and Ballesteros *et al.,* (235) that serine can only hydrogen with the $i-3$ or $i-4$ carbonyl oxygen atom of the same helix. United atomic MD simulations presented a model of *inter*helical serine side chain hydrogen bonding in the presence of a small crossing angle, relative to an AZ2 control. The drop in crossing angle may also lead to an increase in enthalpic contributions at the helix-helix interface. This would fit with the increase in propensity for self-association of AZ2L6S and AZ2L10S according to the GALLEX assay results.

8.2 Limitations

During this study, efforts were made to overcome experimental and computational limitations. What follows is an account of these limitations and where possible, the measures taken.

Firstly, calculating the free energy of oligomerisation from an IPTG titration curve assumes a simplified equilibrium model, yet in truth, there may be any number of contributions from equilibria of molecules native to the cellular environment. An additional complication arises when scaling raw β-galacotosidase measurements by protein expression values using immunoblots. Not only is it impossible to load every cell sample onto the same gel, but also loading a control across multiple gels may prove to lead to technical inaccuracies given that multiple lanes of the same control sample on the same gel yielded statistically different band intensities. Alternatively, one could use multiple controls on one gel, and then take an average. The average would then be used to scale the few remaining cell sample bands across multiple gels. Fortunately, by blotting for the MBP region of the fusion protein, the native MBP of SU101 is also visualised, which is independent of IPTG induction. Assuming that cell growth conditions were approximately the same, the average of the native MBP band intensities across a whole gel could be used as a scaling factor across multiple

gels. Further more, the pixel-intensity values of protein bands is extremely sensitive according to ImageJ measurements. The slightest inaccuracy in manually selecting each band would result in significantly large changes to the β-galactosidase measurement after scaling. As a compromise, expression levels were adjusted to a log-base, to keep visible variations in intensity present but to minimise error.

The PMF calculations of the native sequence Neu and the oncogenic form revealed long range-lipid mediated association. Umbrella windows well beyond the length of a reaction coordinate initially expected were required, totalling 320 simulations at 40 ns per window for a total simulation time of 12.8 μs. These calculations do not include the initial equilibrium time for peptide insertion and bilayer adjustment. In addition, only half of the data points were used to construct each umbrella window histogram distribution, as membrane simulations are notorious for having long auto-correlation times (198). A further limitation to PMF calculations was the obvious difference between the point of free energy convergence on a PMF profile between a united atom force field system to a CG system. These were discussed in length in chapter 6.

Finally, dimer conformational searches were performed using the CHI algorithm. This toolkit provides a quick and computationally inexpensive means of sampling the potential energy landscape between TM helical domains in a user specified dielectric implicit environment. Yet the algorithm does not take into account the variation in the dielectric value across the bilayer, or environmental contributions from hydrophobic mismatch. Yet the CHI algorithm provides an indispensable means of reporting side chain interaction energies.

8.3 Future work

Having investigated the free energy of association of proto- and oncogenic Neu TM domains along a together-apart reaction coordinate, we think that we have only began to capture the complexity behind the dynamics of the receptor tyrosine kinase TM domain. A repetition of the free energy calculations in a mixed model bilayer would help confirm the dimeric structure as a consequence to hydrophobic mismatch. In addition, an investigation using meta-dynamics of a CG model may reveal free energy minima as a function of helix rotation and crossing angle. This would elicit dynamics which may support or otherwise contradict the idea that families of RTKs activate upon rotation (179, 183, 178).

Having investigated the structure and dynamics of low complexity scaffolds with interaction motifs, we believe that using meta-dynamics could extend the free energy captured in this study further. By converting CG dimer structures acquired at the global free energy minima from this study of the AZ2, AZ2$(GG4)_L$ and AZ2$(GG4)_A$ models into a united atom model, meta-dynamics could provide a comprehensive exploration of a free energy landscape of helix-helix crossing angle. Our results would therefore be less at the mercy of whether we have sufficiently sampled different conformations along an *inter*helical reaction coordinate, when in truth we are very interested in the free energy of different helix-helix interfaces. One of our major results in this chapter was the switch between a left-handed AZ2 dimer to a right-handed dimer after the addition of GG4. This would benefit further investigation in at least two ways. Firstly, a free energy calculation of hetero-association between the combinations of AZ2, AZ2$(GG4)_L$ and AZ2$(GG4)_A$, and secondly investigating the clustering of a mixed peptide system of the aforementioned sequences. The results of which could provide initial thermodynamic results for an early framework of a *de novo* TM peptide switch.

There currently exists several TM protein databases (257, 258, 259, 7, 260), but as yet, a database based on *de novo* designed TM helical proteins does not exist. The closest to a database of protein interaction rules would be the repository of soluble coiled-coils, CC+ (261), which has been identified using the program SOCKET (262). The rules of TM domain engagement presented in this study can form the basis for further studies with the aim of compiling a database of TM domain interaction rules and dynamics.

Bibliography

[1] E. Wallin and G. von Heijne. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, 7(4):1029–1038, 1998.

[2] M. M. Sperotto, S. May, and A. Baumgaertner. A modelling of proteins in membranes. *Chem. Phys. Lipids*, 141:2–29, 2006.

[3] `http://blanco.biomol.uci.edu/mpstruc/`.

[4] S.H. White, A.S. Ladokhin, S. Jayasinghe, and K. Hristova. How membranes shape protein structure. *J. Biol. Chem.*, 276(35):32395–32398, 2001.

[5] J.D. Lear, J.P. Schneider, P. K. Kienker, and W. F. DeGrado. Electrostatic effects on ion selectivity and rectification in designed ion channel peptides. *J. Am. Chem. Soc.*, 119:3212–3217, 1997.

[6] J. M. Johnston, G. A. Cool, J. M. Tomich, and M. S. P. Sansom. Conformation and environment of channel-forming peptides: a simulation study. *Biophys J.*, 90(6):1855–1864, 2006.

[7] S.H. White and W.C. Wimley. Membrane protein folding and stability: physical principles. *Ann. Rev. Biophys. Biomol. Struct.*, 28:319–365, 1999.

[8] H. Wang, L. Barreyro, D. Provasi, I. Djemil, C. Torres-Arancivia, M. Filizola, and I. Ubarretxena-Belandia. Molecular determinants and thermodynamics of the amyloid precursor protein transmembrane domain implicated in alzheimer's disease. *J. Mol. Biol.*, 408(5):879–895, 2011.

[9] R.B. Sutton, D. Fasshauer, R. Jahn, and A.T. Brünger. Crystal structure of a snare complex involved in synaptic exocytosis at 2.4 åresolution. *Nature*, 395:347–353, 1998.

[10] James U. Bowie. Helix packing in membrane proteins. *J. Mol. Biol.*, 272:780–789, 1997.

[11] W.C. Wimley. The versatile beta-barrel membrane protein. *Curr. Opin. Struct. Biol.*, 13(4):404–411, 2003.

[12] L.K. Tamm, H. Hony, and B. Liang. Folding and assembly of β-barrel membrane proteins. *Biochim. Biophys. Acta*, 1666:250–263, 2004.

[13] J.L Popot and D.M. Engelman. Membrane protein folding and oligomerization: The two-stage model. *Biochemistry*, 17:4031–4037, 1990.

[14] D.M. Engelman, Y. Chen, C.N. Chin, Curran A.R., A.M. Dixon, A.D. Dupuy, A.S. Lee, U. Lehnert, E.E. Matthews, Y.K. Reshetnyak, A. Senes, and J.L. Popot. Membrane protein folding: beyond the two stage model. *FEBS. Lett.*, 555(1):122–125, 2003.

[15] S. Palchevskyy, S, O. Posokhoc, Y, B. Olivier, J-L. Popot, B. Pucci, and S. Ladokhin, A. Chaperoning of Insertion of Membrane Proteins into LipidBilayers by Hemifluorinated Surfactants: Application to Diphtheria Toxin. *Biochemistry*, 45:2629–2635, 2006.

[16] T.A. Cross, M. Sharma, M. Yi, and H.-X. Zhou. Influence of solubilizing environments on membrane protein structures. *Trends Biochem. Sci*, 36:117–125, 2011.

[17] Natalie Bordag and Sandro Keller. α-Helical transmembrane peptides: A "Divide and Conquer" approach to membrane proteins. *Chem. Phys. Lipids*, 163:1–26, 2010.

[18] H.-H. Shen, T. Lithgow, and L.L Martin. Reconstitution of membrane proteins into model membranes: Seeking better ways to retain protein activities. *Int. J. Mol. Sci*, 14:1589–1607, 2013.

[19] G. Khelashvili, M.V. LeVine, L. Shi, M. Quick, J.A. Javitch, and H. Weinstein. The membrane protein leut in micellar systems: Aggregation dynamics and detergent binding to the s2 site. *J. Am. Chem. Soc.*, 135:14266–14275, 2013.

[20] H. Brockman. Lipid monolayers: Why use half a membrane to characterize protein-membrain interactions? *Curr. Opin. Struc. Biol.*, 9:438–443, 1999.

[21] J. Seelig. Titration calorimetry of lipid-peptide interactions. *Biochim. Biophys. Acta*, 1331(1):103–116, 1997.

[22] Y. Iwahashi and T. Nakamura. Orientation and reactivity of nadh kinase in proteoliposomes. *J. Biochem.*, 105:922–926, 1989.

[23] H. Kolmar, F. Hennecke, K. Gotze, B. Janzer, B. Vogt, F. Mayer, and H.J. Fritz. Membrane insertion of the bacterial signal transduction protein toxr and requirements of transcription activation studied by modular replacement of different protein substructures. *EMBO J.*, 14(16):3895–530, 1995.

[24] B. Brosig and D. Langosch. The dimerization motif of the glycophorin a transmembrane segment in membranes: Importance of glycine residues. *Protein Sci.*, 7:1052–1056, 1998.

[25] W.P. Russ and D.M. Engelman. Toxcat: A measure of transmembrane helix association in a biological membrane. *Proc. Natl. Acad. Sci. U.S.A.*, 96:863–868, 1999.

[26] D. Schneider and D.M. Engelman. Gallex, a measurement of heterologous association of transmembrane helices in a biological membrane. *J. Biol. Chem.*, 278(5):3105–3111, 2003.

[27] E. Linder, S. Unterreitmeier, A.N.J.A. Ridder, and D. Langosch. An extended toxr possyccat system for positive and negative selection of self-interacting transmembrane domains. *J. Microbiol. Methods*, 69:298–305, 2007.

[28] A. Senes, M. Gerstein, and D.M. Engelman. Statistical analysis of amino acid patterns in transmembrane helices: The gxxxg motif occurs frequently and in association with β-branched residues at neighbouring positions. *J. Mol. Biol.*, 296:921–936, 2000.

[29] Kleiger. G., R. Grothe, P. Mallick, and D. Eisenberg. Gxxxg and axxxa: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry*, 41(19):5990–5997, 2002.

[30] K. R. MacKenzie, J. H. Prestegard, , and D. M. Engelman. A transmembrane helix dimer: structure and implications. *Science*, 276(5309):131–133, 1997.

[31] M.A. Lemmon, J.M. Flanagan, H.R. Treutlein, J. Zhang, and D.M. Engelman. Sequence

specificity in the dimerization of transmembrane alpha-helices. *Biochemistry*, 31:12719–12725, 1992.

[32] P.D. Adams, D.M. Engelman, and A.T. Brünger. Improved prediction for the structure of the dimeric transmembrane domain of glycophorin a obtained through global searching. *Proteins*, 26:257–261, 1996.

[33] C. Finger, T. Volkmer, A. Prodöhl, D.E. Otzen, D.M. Engelman, and D. Schneider. The stability of transmembrane helix interactions measured in a biological membrane. *J. Mol. Biol.*, 358(5):1221–1228, 2006.

[34] K.A. Williams, M. Gilibowicka, Z. Li, H. Li, A.R. Khan, Y.M.Y. Chen, J. Wang, D.A. Marvin, and C.M. Deber. Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage m13: Role of small residues in protein oligomerization. *J. Mol. Biol.*, 252:6–14, 1995.

[35] C.K. Asundi and D.J. Carey. Self-association of n-syndecan (syndecan-3) core protein is mediated by a novel structural motif in the transmembrane domain and ectodomain flanking region. *J. Biol. Chem.*, 270:26404–26410, 1995.

[36] T.E. Herbert, S. Moffet, J-P. Morello, T.P. Loisel, D.G. Bichet, C. Barret, and M. Bouvier. A peptide derived from a beta$_2$-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J. Biol. Chem.*, 271:16384–16392, 1996.

[37] Y.J. Lin, J.G. Peng, and S.C. Wu. Characterization of the gxxxg motif in the first transmembrane segment of japanese encephalitis virus precursor membrane (prm) protein. *J. Biomed. Sci*, 17, 2010.

[38] L. Gaidukov, A.R. Nagel, S.Z. Xu, M. Penman, and M. Krieger. Glycine dimerization motif in the n-terminal transmembrane domain of the high density lipoprotein receptor sr-bi required for normal receptor oligomerization and lipid transport. *J. Biol. Chem.*, 286:18452–18464, 2011.

[39] E.V. Bocharov, K.S. Mineev, P.E. Volynsky, Y.S. Ermolyuk, E.N. Tkach, A.G. Sobol, V.V. Chupin, M.P. Kirpichnikov, R.G. Efremov, and A.S. Arseniev. Spatial structure of the dimeric transmembrane domain of the growth factor of the dimeric transmembrane domain of the

growth factor receptor erbb2 presumably corresponding to the receptor active state. *J. Biol. Chem.*, 283:6950–6956, 2008.

[40] E.V. Bocharov, M.L. Mayzel, P.E. Volynsky, Y.S. Ermolyuk, A.A. Schulga, E.O. Artemenko, R.G. Efremov, and A.S. Arseniev. Spatial structure and ph-dependent conformation diversity of dimeric transmembrane domain of the receptor tyrosine kinase epha1. *J. Biol. Chem.*, 283:29385–29395, 2008.

[41] E.V. Bocharov, Y.E. Pustovalova, K.V. Pavlov, P.E. Volynsky, M.V. Goncharuk, Y.S. Ermolyuk, D.V. Karpunin, A.A. Schulga, M.P. Kirpichnikov, R.G. Efremov, I.V. Maslennikov, and A.S. Arseniev. Unique dimeric structure of bnip3 transmembrane domain suggests membrane permeabilization as a cell death trigger. *J. Biol. Chem.*, 282:16256–16266, 2007.

[42] E.S. Sulistijo and K.R. MacKenzie. Structural basis for dimerization of the bnip3 transmembrane domain. *Biochemistry*, 48:5106–5120, 2009.

[43] T.L. Lau, C. Kin, M.H. Ginsberg, and T.S. Ulmer. The structure of the integrin alpha iib beta 3 transmembrane complex explains integrin transmembrane signalling. *EMBO J.*, 28:1351–1261, 2009.

[44] J. Yang, Y.Q. Ma, R.C. Page, S. Misra, E.F. Plow, and J. Qin. Structure of an integrin alpha iib beta 3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proc. Natl. Acad. Sci. U.S.A.*, 106:17729–17734, 2009.

[45] K.S. Mineev, E.V. Bocharov, Y.E. Pustovalova, O.V. Bocharova, V.V. Chupin, and A.S. Arseniev. Spatial structure of the transmembrane domain heterodimer of erbb1 and erbb2 receptor tyrosine kinase. *J. Mol. Biol.*, 400:231–243, 2010.

[46] A. Beevers, A. Nash, M. Salazar-Cancino, D. Scott, R. Notman, and A. Dixon. Effects of the oncogenic v_{664}e mutation on membrane insertion, structure, and sequence-dependent interactions of the neu transmembrane domain in micelles and model membranes: an integrated biophysical and simulation study. *Biochemistry*, 51(12):2558–2568, 2012.

[47] A.K Doura, F.J. Kobus, L. Dubrovsky, E. Hibbard, and K.G. Fleming. Sequence context modulates the stability of a gxxxg-mediated transmembrane helix-helix dimer. *J. Mol. Biol.*, 341:991–998, 2004.

[48] A.K. Doura and K.G. Fleming. Complex interactions at the helix-helix interface stabilize the glycophorin a transmembrane dimer. *J. Mol. Biol.*, 343:1487–1497, 2004.

[49] S. Unterreitmeier, A. Fuchs, T. Schaffler, R.G. Heym, G. Frishman, and D. Langosch. Phenylalanine promotes interaction of transmembrane domains via gxxxg motifs. *J. Mol. Biol.*, 374:705–718, 2007.

[50] J.R. Herrmann, J.C. Panitz, S. Unterreitmeier, A. Fuchs, D. Frishman, and L. Langosch. Complex patterns of histodine, hydroxylated amino acids and the gxxxg motif mediate high-affinity transmembrane domain interactions. *J. Mol. Biol.*, 385:912–923, 2009.

[51] M. Orzáez, D. Lukovic, C. Abad, E. Pérez-Payá, and I. Mingarro. Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin a dimerisation motif. *FEBS Lett.*, 579:1633–1638, 2005.

[52] A.D. McLachlan and M Steward. Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. *J. Mol. Biol.*, 98:293–304, 1975.

[53] E.K. O'Shea, J.D. Klemm, P.S. Kim, and T. Alber. X-ray structure of the gcn4 leucine zipper, a two stranded, parallel coiled coil. *Science*, 243:539–544, 1991.

[54] D. Langosch and J. Heringa. Interaction of transmembrane helices by a knobs-into-holes geometry characteristic of soluble coiled coils. *Proteins: Struct., Funct., Genet.*, 31:150–160, 1998.

[55] `http://what-when-how.com/molecular-biology/heptad-repeat-molecular-biology/`.

[56] A.W. Roszak, V. Moulisova, A.D. Reksodipuro, A.T. Gardiner, R. Fujii, H. Hashimoto, N.W. Isaacs, and Cogdell. R.J. New insights into the structure of the reaction centre from blastochloris viridis: evolution in the laboratory. *Biochem. J.*, 442:27–37, 2012.

[57] J. Deisenhofer, O. Epp, K. Miki, R. Juber, and H. Michel. Structure of the photosynthetic reaction center from the purple bacterium *Rhodopseudomonas virides*. *Nature*, 318:618–623, 1985.

[58] R. Henderson, J.M. Baldwin, T.A. Ceska, F. Zemlin, Beckmannm E., and K.H. Downing.

Model for the structure of bacteriorhodopsin based on high resolution electron cryomicroscopy. *J. Mol. Biol.*, 213:899–911, 1990.

[59] S. Iwata, C. Ostermeier, B. Ludwig, and H. Michel. Structure at 2.8 å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature*, 376:660–669, 1995.

[60] R. Gurezka, R. Laage, B. Brosig, and D. Langosch. A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J. Biol. Chem.*, 274:9265–9270, 1990.

[61] G. von Heijne. Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.*, 7(12):909–918, 2006.

[62] E. Granseth, G. von Heijne, and A. Elofsson. A study of the membrane-water interface region of membrane proteins. *J. Mol. Biol.*, 361:591–603, 2006.

[63] W.C. Winley, T.P. Creamer, and S.H White. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry*, 35(16):5109–5124, 1996.

[64] W.C. Wimley and S.H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.*, 3:842–848, 1996.

[65] D. M. Engelman, T. A. Steitz, and A. Goldman. Identifying non-polar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biomol. Struct.*, 15:321–353, 1986.

[66] T. Hessa, N. Meindl-Beinker, A. Bernsel, J. Kim, Y. Sato, M. Lerch, C. Lundin, I. Nilsson, S.H. White, and G. von Heijne. Molecular code for transmembrane-helix recognition by the sec61 translocon. *Nature*, 450:1026–1030, 2007.

[67] C. Choma, H. Gratkowski, J.D. Lear, and W.F. DeGrado. Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.*, 7(2):161–166, 2000.

[68] E.K. O'She, J.D. Klemm, P.S. Kim, and T.A. Alber. X-ray structure of the gcn4 leucine zipper, a two-stranded coiled coil. *Science*, 254:539–544, 1991.

[69] H. Gratkowski, J.D. Lear, and W.F. DeGrado. Polar side chains drive the association of polyleucine transmembrane helices. *Proc. Natl. Acad. Sci. U.S.A.*, 98(5):880–885, 2001.

[70] F.X. Zhou, H.J. Merianos, A.T. Brunger, and D.M. Engelman. Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl. Acad. Sci. U.S.A.*, 85(5):2250–2255, 2001.

[71] J.P. Dawson, J.S. Weinger, and D.M. Engelman. Motifs of serine and threonine can drive association of transmembrane helices. *J. Mol. Biol.*, 316:799–805, 2002.

[72] W. J. Gullick and M. J. Sternberg. Neu receptor dimerization. *Nature*, 339:587, 1989.

[73] W. J. Gullick, A. C. Bottomley, F. J. Lofts, D. G. Doak, D. Mulvey, R. Newman, M. J. Crumpton, M. J. E. Sternberg, and I. D. Campbell. Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein. *EMBO J.*, 11(1):43–48, 1992.

[74] W. J. Gullick, , and M. J Sternberg. A sequence motif in the transmembrane region of growth factor receptors with tyrosine kinase activity mediates dimerization. *Protein Eng.*, 3(4):245–248, 1990.

[75] J.D. Lear, H. Gratkowski, L. Adamian, J. Liang, and W.F. DeGrado. Position-dependence of stabilizing polar interactions of asparagine in transmembrane helical bundles. *Biochemistry*, 42:6400–6407, 2003.

[76] C. Baeza-Delgado, M.A. Marti-Renom, and I. Mingarro. Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.*, 42:199–207, 2012.

[77] H.J. Sharpe, T.J. Stevens, and S. Munro. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell*, 142:158–169, 2010.

[78] J.P. Dawson, R.A. Melnyk, C.M. Deber, and D.M. Engelman. Sequence context strongly modulates association of the polar residues in transmembrane helices. *J. Mol. Biol.*, 331:255–262, 2003.

[79] J.R. Herrmann, A. Fuchs, J.C. Panitz, T. Eckert, S. Unterreitmeier, D. Frishman, and D. Langosch. Ionic interactions promote transmembrane helix-helix association depending on sequence context. *J. Mol. Biol.*, 392(2):452–461, 2010.

[80] D.V Tulumello and C.M. Deber. Efficiency of detergents at maintaining membrane protein structures in their biological relevant forms. *Biochim. Biophys. Acta*, 1818:1351–1358, 2012.

[81] H. Hong and J.U Bowie. Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J. Am. Chem. Soc.*, 133:11389–11398, 2011.

[82] C.R. Raetz. Enzymology, genetics, and regulation of membrane phospholipid synthesis in escherichia coli. *Microbiol. Rev*, 42:614–659, 1978.

[83] O.S. Andersen and R.E. Koeppe II. Bilayer thickness and membrane protein function: An energetic perspective. *Annu. Rev. Biophys. Biomol. Struct.*, 36:107–130, 2007.

[84] E. V. Bocharov, M. L. Mayzel, P. E. Volynsky, K. S. Mineev, E. N. Tkach, Y. S. Ermolyuk, A. A. Schulga, R. G. Efremov, and Arseniev A. S. Left-handed dimer of epha2 transmembrane domain: helix packing diversity among receptor tyrosine kinases. *Biophys J.*, 98:881–889, 2010.

[85] C. Muhle-Goll, S. Hoffmann, S. Afonin, S.L. Grage, A.A. Polyansky, D. Windisch, M. Zeitler, J. Bürck, and A.S. Ulrich. Hydrophobic matching controls the tilt and stability of the dimeric platelet-derived growth factor receptor (pdgfr) β transmembrane segment. *J. Biol. Chem.*, 287(31):26178–26186, 2012.

[86] S. Ozdirekcan, D.T. Rijkers, R.M. Liskamp, and J.A. Killian. Influence of flanking residues on tilt and rotation angles of transmembrane peptides in lipid bilayer. a solid-state ^{2}h nmr study. *Biochemistry*, 44:1004–1012, 2005.

[87] J. Ren, S. Lew, J. Wang, and E. London. Control of transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. *Biochemistry*, 38:5905–5912, 1999.

[88] S. Mall, R. Broadbridge, R.P. Sharma, J.M. East, and A.G. Lee. Self-association of model transmembrane alpha-helices is modulated by lipid structure. *Biochemistry*, 40:12379–12386, 2001.

[89] E. Sparr, W.L. Ash, P.V. Nazarov, D.T.S. Rijkers, M.A. Hemminga, D.P. Tieleman, and J.A. Killlian. Self-association of transmembrane α-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch. *J. Biol. Chem.*, 280:39324–39331, 2005.

[90] H.I. Petrache, G.M. Zuckerman, J.N. Sachs, J.A. Killian, R.E. Koeppe II, and T.B. Woolf. Hydrophobic matching mechanism investigated by molecular dynamics. *Langmuir*, 18:1340–1351, 2002.

[91] M. Yiannourakou, L. Marsella, F. de Meyer, and B. Smit. Towards an understanding of membrane-mediated protein-protein interactions. *Faraday Discuss.*, 144:359–367, 2010.

[92] de Planque M.R.R., J.A.W. Kruijtzer, R.M.J. Liskamp, D. Marsh, D.V. Greathouse, R.E. Koeppe, B. de Kruijff, and J.A. Killian. Different membrane anchoring positions of tryptophan and lysine in synthetic transmembrane α-helical peptides. *J. Biol. Chem.*, 274:20839–20846, 1999.

[93] E. Strandberg, S. Morein, D.T.S. Rijkers, P.C.A. van der Wel, R.M.J. Liskamp, and J.A. Killian. Lipid dependence of membrane anchoring properties and snorkeling behaviour of aromatic and charged residues in transmembrane peptides. *Biochemistry*, 41:7190–7198, 2002.

[94] D.P. Siegel, V. Cherezov, D.V. Greathouse, R.E. Koeppe, J.A. Killian, and M. Caffrey. Transmembrane peptides stabalize inverted cubic phases in a biphasic length-dependent manner: implications for protein-induced membrane fusion. *Biochemistry*, 90:200–211, 2006.

[95] R.J. Webb, J.M. East, R.P. Sharma, and A.G. Lee. Hydrophobic mismatch and the incorporation of peptides in lipid bilayers: a possible mechanics for retention in the golgi. *Biochemistry*, 37:673–679, 1998.

[96] A.A. Polyansky, P.E. Volynsky, and G.E. Efremov. Multistate organization of transmembrane helical protein dimers governed by the host membrane. *J. Am. Chem. Soc.*, 134(35):14390–14400, 2012.

[97] W.P. Russ and D.M. Engelman. The gxxxg motif: A framework for transmembrane helix-helix association. *J. Mol. Biol.*, 296:911–919, 2000.

[98] L. Ruan and D. Langosch. The interface of a membrane-spanning leucine zipper mapped by asparagine-scanning mutagenesis. *Protein Sci.*, 13:555–559, 2004.

[99] N. Sal-man, D. Gerber, and Y. Shai. The identification of a minimal dimerization motif qxxs that enables homo- and hetro-association of transmembrane helices *in vivo*. *J. Biol. Chem.*, 280(29):27449–27457, 2005.

[100] J. Lee and W. Im. Role of hydrogen bonding and helix-lipid interactions in transmembrane helix association. *J. Am. Chem. Soc.*, 130:6456–6462, 2008.

[101] Z. Shi, C.A. Olson, A.G. Bell, and N.R. Kallenbach. Non-classical helix-stabilizing interactions: C-h· · ·o-h bonding between phe and glu side chains in alpha-helical peptides. *Biophys. Chem.*, 101–102:267–279, 2002.

[102] S.A. Jusoh and V. Helms. Helical integrity and microsolvation of transmembrane domains from flaviviridae envelope glycoproteins. *Biochim. Biophys. Acta*, 2011:1040–1049, 2011.

[103] R.M. Johnson, K. Hecht, and C.M. Deber. Aromatic and cation-pi interactions enhance helix-helix association in a membrane environment. *Biochemistry*, 46:9208–9214, 2007.

[104] A. Ridder, P. Skupjen, S. Unterreitmeier, and D. Langosch. Tryptophan supports interaction of transmembrane helices. *J. Mol. Biol.*, 354:894–902, 2005.

[105] N. Sal-man, D. Gerber, I. Bloch, and Y. Shai. Specificity in transmembrane helix-helix interactions mediated by aromatic residues. *J. Biol. Chem.*, 282(27):19753–19761, 2007.

[106] R. Gurezka, R. Laage, B. Brosig, and D. Langosch. A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J. Biol. Chem.*, 274(14):9265–9270, 1999.

[107] M.W. Hofmann, K. Weise, J. Ollesch, P. Agrawal, H. Stalz, W. Stelzer, F. Hulsbergen, H. de Groot, K. Gerwert, J. Reed, and D. Langosch. De novo design of conformationally flexible transmembrane peptides driving membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.*, 101(41):14776–14781, 2004.

[108] E.H.C. Bromley, K. Channon, E. Moutevelis, and D.N. Woolfson. Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *ACS Chem. Biol.*, 3:38–50, 2008.

[109] K. Channon, E.H.C. Bromley, and D.N. Woolfson. Synthetic biology through biomolecular design and engineering. *Curr. Opin. Struct. Biol.*, 18:491–498, 2008.

[110] I. V. Korendovych, A. Senes, Y. H. Kim, J. D. Lear, C. Fry, M. J. Therien, J. K. Blasie, F. A. Walker, and W. F. DeGrado. De Novo Design and Molecular Assembly of a Transmembrane Diphorphyrin-Binding Protein Complex. *J. Am. Chem. Soc.*, 132:15516–15518, 2010.

[111] J. M. Cordova, P. L. Noack, S. A. Holcove, J. D. Lear, and G. Ghirlanda. Design of a functional membrane protein by engineering a heme-binding site in glycophorin A. *J. Am. Chem. Soc*, 129(3):512–518, 2007.

[112] S. Shinde, J. M. Cordova, B. W. Woodrum, and G. Ghirlanda. Modulation of function in a minimalist heme-binding membrane protein. *J. Biol. Inorg. Chem.*, 17(4):557–564, 2012.

[113] R. S. Signarvic and W. F. DeGrado. Metal-Binding Dependent Disruption of Membranes by Designed Helices. *J. Am. Chem. Soc.*, 131:3377–3384, 2009.

[114] N. G. Tew, D. Liu, B. Chen, R. J. Doerksen, J. Kaplan, P. J. Carroll, M. L. Klein, and W. F. DeGrado. *De novo* design of biomimetic antimicrobial polymers. *PNAS*, 99(8):5110–5114, 2002.

[115] R. W. Scott, W. F. DeGrado, and G. N. Tew. De Novo Designed Synthetic Mimics of Antimicrobial Peptides. *Curr. Opin. Biotechnol.*, 19(6):620–627, 2008.

[116] T. Miyamoto, S. Razavi, R. DeRose, and T. Inoue. Synthesizing biomolecule-based boolean logic gates. *ACS Synth. Biol.*, 2:72–82, 2013.

[117] A.P. de Silva and S. Uchiyama. Molecular logic gates and luminescent sensors based on photoinduced electron transfer. *Top. Curr. Chem.*, 300:1–28, 2011.

[118] S.S. Shekhawat, J.R. Porter, A. Sriprasad, and I. Ghosh. An autoinhibited coiled-coil design strategy for split-protein protease sensors. *J. Am. Chem. Soc.*, 131(42):15284–15290, 2009.

[119] W. de Ronde, P. Rein ten Wolde, and A. Mugler. Protein logic: a statistical mechanical study of signal integration at the single-molecule level. *Biophys. J.*, 103(5):1097–1107, 2012.

[120] R. O'Brien, J.E. Ladbury, and B.Z. Chowdry. *Protein-Ligand interactions: hydrodynamics and calorimetry Edition*. 2000.

[121] P. Atkins and J. de Paula. *Elements of Physical Chemistry*, volume 4th Eddition. Oxford University Press, 2005.

[122] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 manual and user guide.* Verlag der Fachvereine, 1996.

[123] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, P.D. Tieleman, and S-J. Marrink. The martini coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.*, 4:819–834, 2008.

[124] J. E. Lennard-Jones. On the determination of molecular fields. *P. Roy. Soc. Lond. A Mat.*, 106:463–477, 1924.

[125] R. A. Buckingham. The classical equation of state of gaseous helium, neon and argon. *P. Roy. Soc. Lond. A Mat.*, 168:264–283, 1938.

[126] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.*, 64:253–287, 1921.

[127] T. A. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[128] U. Essman, L. Perela, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8592, 1995.

[129] W. F. van Gunsteren and H. J. C. Berendsen. *Gromos-87 manual*. 1987.

[130] P.M. Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Phys. Rev.*, 34:57–64, 1929.

[131] D. M. Ferguson. Parametrization and evaluation of a flexible water model. *J. Comput. Chem.*, 16:501–511, 1995.

[132] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, and M. Swaminathan, S. Karplus. Charmm: a program for macromolecular energy, a minimization, and dynamics calculation. *J. Comput. Chem.*, 4:187–217, 1983.

[133] C. P. Lawrence and J. L. Skinner. Flexible tip4p model for molecular dynamics simulation of liquid water. *Chem. Phys. Lett.*, 372:842–847, 2003.

[134] J. P. Ryckaert and A. Bellemans. Molecular dynamics of liquid alkanes. *J. Chem. Soc. Faraday Trans.*, 66:95–106, 1978.

[135] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.*, 102:6665–6670, 2005.

[136] H.J. Risselada and S.J. Marrink. Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, 11:2056–2067, 2009.

[137] M. G. Saunders and G. A. Voth. Comparison between actin filament models: Coarse-graining reveals essential differences. *Structure*, 20:641–653, 2012.

[138] M. Levitt and A. Warshel. Computer simulation of protein folding. *Nature*, 253:694–698, 1975.

[139] Y. Ueda, H. Taketomi, and N. Gō. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, 17:1531–1548, 1978.

[140] A. Liwo, S. Oldziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H.A. Scheraga. A united-residue force field for off-lattice protein-structure simulations. 1. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comp. Chem*, 18:849–873, 1997.

[141] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. A coarse grain model for phospholipid simulations. *J. Phys. Chem. B*, 105:4464–4470, 2001.

[142] J.A. Snyman. *Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms*. Springer Publishing, 2005.

[143] K. Zimmerman. All purpose molecular mechanics simulator and energy minimizer. *J. Comput. Chem.*, 12:310–319, 1991.

[144] C.G. Broyden. The convergence of a class of double-rank minimisation algorithms. *JIMA*, 6:76–90, 1970.

[145] R. Fletcher. A new approach to variable metric algorithms. *Comput. J*, 13(3):317–322, 1970.

[146] D. J. Adams, E. M. Adams, and G. J. Hills. The computer simulation of polar liquids. *Mol. Phys.*, 38:387–400, 1979.

[147] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity-rescaling. *J. Chem. Phys*, 126:014101, 2007.

[148] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[149] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNoal, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684, 1984.

[150] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[151] G. S. Grest and K. Kremer. Molecular-dynamics simulation for polymers in the presence of a head bath. *Phys. Rev. A*, 33(5):3628–3631, 1986.

[152] M. Ceriotti, G. Bussi, and M. Parrinello. Colored-Noise Thermostats a Ìla Carte. *J. Chem. Theory Comput.*, 6(4):1170–1180, 2010.

[153] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7191, 1981.

[154] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *J. Chem. Phys.*, 23:327–341, 1977.

[155] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463–1472, 1997.

[156] W. Yang. Special issue: Free energy simulations: Methods and applications. *J. Comput. Chem.*, 30(11):1615–1747, 2009.

[157] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[158] G. M. Torrie and J. P. Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chem. Phys. Lett.*, 28(4):578–581, 1974.

[159] S Kumar, J. M. Rosenberg, B Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[160] B. M. Dickson. Approaching a parameter-free metadynamics. *Phys. Rev. E*, 84(3):37701–37705, 2011.

[161] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, 99(22):12562–12566, 2002.

[162] J. Hénin, A. Pohorille, and C. Chipot. Insights into the Recognition and Association of Transmembrane α-Helices. The Free Energy of α-Helix Dimerization in Glycophorin A. *J. Am. Chem. Soc.*, 127:8478–8484, 2005.

[163] J. Shen and J.A McCammon. Molecular dynamics simulation of superoxide interacting with superoxide dismutase. *Chem. Phys.*, 158(2–3):191 – 198, 1991.

[164] J. Kästner and W. Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: 'umbrella integration'. *J. Chem. Phys.*, 123:144104(1–5), 2005.

[165] A. M. Ferrenberg and R. H. Swendsen. Optimized monte carlo data analysis. *Phys. Rev. Lett.*, 63(12):1095–1098, 1989.

[166] C.H. Bennett. Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys.*, 22(2):245–268, 1976.

[167] J. Sambrook and D. W. Russell. *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory Press, 2001.

[168] N. A. Treptow and H. A. Shuman. Genetic evidence for substrate and periplasmic-binding-protein recognition by the malf and malg protein, sytoplasmic membrane components of the escherichia coli maltose transport system. *J. Bacteriol.*, 163:654–660, 1985.

[169] M. Dmitrova, G. Younès-Cauet, P. Oertel-Buchheit, D. Porte, M. Schnarr, and M. Granger-Schnarr. A new lexa-based genetic system for monitoring and analyzing protein heterodimerization in escherichia coli. *Mol. Gen. Genet.*, 257(2):205–212, 1998.

[170] W.S. Rasband. ImageJ (software), 1997-2013.

[171] J. R. Broughman, L. P. Shank, O. Prakash, B.D. Schultz, T. Iwamoto, J. M. Tomich, and K.E. Mitchell. Structural implications of placing cationic residues at either the nh_2- or cooh-terminus in a pore-forming synthetic peptide. *J. Membr. Biol.*, 190(2):93–103, 2002.

[172] G. A. Cook, O. Prakash, K. Zhang, L. P. Shank, W. A. Takeguchi, A. S. Robbins, Y. X. Gong, T. Iwamoto, B. D. Schultz, and J. M. Tomich. Activity and structural comparisons of solution associating and monomeric channel-forming peptides derived from the glycine receptor m2 segment. *Biophys J.*, 86(3):1424–1435, 2004.

[173] P. Rang, P.L. Mandal, and Y. Xu. NMR structures of the second transmembrane domain of the human glycine receptor α_1-subunit: model of pore architecture and channel gating. *Biophys J.*, 274(49):252–262, 2002.

[174] E. London, K. Shahidullah, and S.S. Krishnakumar. The effect of hydrophilic substitutions and anionic lipids upon the transverse positioning of the transmembrane helix of the erbb2 (neu) protein incorporated into model membrane vesicles. *J. Mol. Biol.*, 396(1):209–220, 2010.

[175] J.M. Mendrola, M.B. Berger, M.C. King, and M.A. Lemmon. The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J. Biol. Chem.*, 277(7):4704–4712, 2002.

[176] K. Hristova, L. He, and N. Shobnam. Specific inhibition of a pathogenic receptor tyrosine kinase by its transmembrane domain. *Biochim. Biophys. Acta*, 1808(1):253–259, 2011.

[177] N. Garnier, M. Genest, P. Aller, and L. Voiry. Molecular dynamics (md) investigations of preformed structures of the transmembrane domain of the oncogenic neu receptor dimer in a dmpc bilayer. *Biopolymers*, 77(4):184–197, 2005.

[178] A.J. Beevers, A. Damianoglou, J. Oates, A. Rodger, and A.M. Dixon. Sequence-dependent oligomerization of the neu transmembrane domain suggests inhibition of "conformational switching" by an oncogenic mutant. *Biochemistry*, 49(13):2811–2820, 2010.

[179] C. A. Bell, J. A. Tynan, K. C. Hart, A. N. Meyer, S. C. Robertson, and D. J. Donoghue. Rotational coupling of the transmembrane and kinase domains of the neu receptor tyrosine kinase. *Mol. Biol. Cell*, 11:3589–3599, 2000.

[180] J Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 103:211–225, 2000.

[181] N. Seubert, Y. Royer, J. Staerk, K. F. Kubatzky, V. Moucadel, S. Krishnakumar, S. O. Smith, , and S. N. Constantinescu. Active and inactive orientations of the transmembrane and cytosolic domains of the eerythropoietin receptor dimer. *Mol. Cell*, 12:1239–1250, 2003.

[182] X. Yu, K. D. Sharma, T. Takahashi, R. Iwamoto, , and E. Mekada. Ligand-independent dimer formation of epidermal growth factor receptor (egfr) is a step separable from ligand-induced egfr signaling. *Mol. Biol. Cell*, 13:2547–2557, 2002.

[183] T. Moriki, H. Maruyama, and I. N. Maruyama. Activation of preformed egf receptor dimers by ligand-induced rotation of the transmembrane domain. *J. Mol. Biol.*, 311:1011–1026, 2001.

[184] S. J. Fleishman, J. Schlessinger, and N. Ben-Tal. A putative molecular-activation switch in the transmembrane domain of erbb2. *Proc. Natl. Acad. Sci. U.S.A.*, 99:15937–15940, 2002.

[185] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. von Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos froce-field paramter sets 53a5 and 53a6. *J. Comput. Chem.*, 25(13):1656–1676, 2004.

[186] Y. Mu, D.S. Kosov, and G.J. Stock. Theory of phase transitions in polypeptides and proteins. *Physical Chemistry B*, 107:5064–50, 2003.

[187] K. Toukan and A. Rahman. Molecular-dynamics study of atomic motions in water. *Phys. Rev. B*, 31:2643–2648, 1985.

[188] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[189] A. Kukol. Lipid models for united-atom molecular dynamics simulations of protein. *J. Chem. Theory Comput.*, 5:615–626, 2009.

[190] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[191] P.D. Adams, I.T. Arkin, D.M. Engelman, and A.T. Brünger. Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nat. Struct. Biol.*, 2:154–162, 1995.

[192] N. Sajot and M. Genest. Structure prediction of the dimeric neu/erbb-2 transmembrane domain from multi-nanosecond molecular dynamics simulations. *Eur. Biophys. J. Biophys. Lett.*, 28:648–662, 2000.

[193] S.O. Smith, C.S. Smith, and B.J. Bormann. Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the neu/erbb-2 receptor. *Nat. Struct. Biol.*, 3:252–258, 1996.

[194] N. Kučerka, M-P. Nieh, and J. Katsaras. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim. Biophys. Acta*, 1808:2761–2771, 2011.

[195] K. Mitra, I. Ubarretxena-Belandia, T. Taguchi, G. Warren, and D.M. Engelman. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc. Natl. Acad. Sci. U.S.A.*, 101(12):4083–4088, 2004.

[196] J. Domański, P. Stansfeld, M.S.P. Sansom, and O. Beckstein. Lipidbook: A public repository for force field parameters used in membrane simulations. *J. Membr. Biol.*, 236:255–258, 2010.

[197] C. Kandt, W.L. Ash, and D.P. Tieleman. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, 41(4):475–488, 2007.

[198] J.S. Hub, B.L. de Groot, and D. van der Spoel. g_wham - a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.*, 6(12):3713–3720, 2010.

[199] W. J. Allen, J. A. Lemkul, and D. R Bevan. Gridmat-md: A grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.*, 30(12):1952–1958, 2009.

[200] M. Melicherčík, A. Holúbeková, T. Hianik, and J Urban. Effect of the aminoacid composition of model α-helical peptides on the physical properties of lipid bilayers and peptide conformation: a molecular dynamics simulation. *J. Mol. Model.*, 19(11):4723–4730, 2013.

[201] W. Humphrey, A. Dalke, and K. Schulten. Vmd - visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

[202] J. A. Killian. Hydrophobic mismatch between proteins and lipids in membranes. *Biochim. Biophys. Acta*, 1376(3):401–415, 1998.

[203] J. A. Killian. Synthetic peptides as models for intrinsic membrane proteins. *FEBS Lett.*, 555(1):134–138, 2003.

[204] M. R. de Planque, B. B. Bonev, J. A. Demmers, D. V. Greathouse, R. E. Koeppe, F. Separovic, A. Watts, and J. A. Killian. Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions. *Biochemistry*, 42:5341–5248, 2003.

[205] S.O. Smith, C.S. Smith, and B.J. Bormann. Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the Neu/erbB-2 receptor. *Nat. Struct. Biol.*, 3(2), 1996.

[206] P.W. Brandt-Rauf, M.R. Pincus, and J.M Chen. Conformational changes induced by the transforming amino acid substitution in the transmembrane domain of the neu oncogene-encoded p185 protein. *J. Protein Chem.*, 8:749–756, 1989.

[207] P.W. Brandt-Rauf, S. Rackovsky, and M.R Pincus. Correlation of the structure of the transmembrane domain of the neu oncogene-encoded p185 protein with its function. *Proc. Natl. Acad. Sci. U.S.A.*, 87:8660–8664, 1990.

[208] R.S. Houliston, R.S. Hodges, F.J. Sharom, and J.H. Davis. Characterization of the proto-oncogenic and mutant forms of the transmembrane region of neu and micelles. *J. Biol. Chem.*, 279(4):24073–20480, 2004.

[209] A. Prakash, L. Janosi, and M. Doxastakis. Gxxxg motifs, phenylalanine, and cholesterol guide the self-association of transmembrane domains of erbb2 receptors. *Biophys. J.*, 101:1949–1958, 2011.

[210] A Grafmuller, R Lipowsky, and V Knecht. Effect of tension and curvature on the chemical potential of lipids in lipid aggregates. *Phys. Chem. Chem. Phys.*, 15:876–881, 2013.

[211] S. S. Krishnakumar and E. London. The control of transmembrane helix transverse position in membranes by hydrophilic residues. *J. Mol. Biol.*, 374:1251–1269, 2007.

[212] C. Loudet, L. Khemtémourian, F. Aussenac, S. Gineste, M.F. Achard, Erick, and J. Dufourc. Bicelle membranes and their use for hydrophobic peptide studies by circular dichroism and solid state nmr. *Biochim. Biophys. Acta*, 1724:315–323, 2005.

[213] D.F. Stern. Tyrosine kinase signalling in breast cancer erbb family receptor tyrosine kinases. *Breast Cancer Res.*, 2:176–183, 2000.

[214] P.J. Brennan, T Kumogai, A. Berezov, R. Murali, and M.I. Greene. HER2/Neu: mechanism of dimerization/oligomerization. *Oncogene*, 19:6093–6101, 2000.

[215] L. He and K. Hristova. Pathogenic activation of receptor tyrosine kinases in mammalian membrane. *J. Mol. Biol.*, 384:1130–1142, 2008.

[216] X. Prasanna, P.J. Praveen, and D. Sengupta. Sequence dependent lipid-mediated effects modulate the dimerization of erbb2 and its associative mutants. *Phys. Chem. Chem. Phys.*, 15:19031–19041, 2013.

[217] K.R. MacKenzie. Folding and stability of alpha-helical integral membrane proteins. *Chem. Rev.*, 106:1931–1977, 2006.

[218] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White, and G. von Heijne. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433:377–381, 2005.

[219] L. P. Liu, S. C. Li, N. K. Goto, and C. M. Deber. Threshold hydrophobicity dictates helical conformations of peptides in membrane environments. *Biopolymers*, 39:465–470, 1999.

[220] B. Bechinger. Membrane insertion and orientation of polyalanine peptides: A ^{15}N solid-state nmr spectroscopy investigation. *Biophys. J.*, 81:2251–2256, 2001.

[221] Y.P. Zhang, N. A. H. L. Ruthven, R. S. Hodges, and R. N. McElhaney. Interaction of a peptide model of a hydrophobic transmembrane α-helical segment of a membrane protein with phosphatidylcholine bilayers: differential scanning calorimetric and ftir spectroscopic studies. *Biochemistry*, 31:11579–11588, 1992.

[222] J Ren, S. Lew, J. Wang, and E. London. Control of the transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. *Biochemistry*, 38(18):5905–5912, 1999.

[223] K. Illergøard, A. Kauko, and A. Elofsson. Why are polar residues within the membrane core evolutionary conserved? *Proteins*, 79:79–91, 2011.

[224] D. Langosch and J. Heringa. Interaction of transmembrane helices by knobs-into-holes packing characteristic of soluble coiled coils. *Proteins: Struct., Funct., Genet.*, 31:150–159, 1998.

[225] J.D. Lear, A.L. Stouffer, H. Gratkowski, V. Nanda, and W.F. DeGrado. Association of a model transmembrane peptide containing gly in a heptad sequence motif. *Biophys J.*, 87:3421–3429, 2004.

[226] N. Sal-Man, D. Gerber, and Y. Shai. The composition rather than position of polar residues (qxxs) drives aspartate receptor transmembrane domain dimerization in vivo. *Biochemistry*, 43(8):2309–2013, 2004.

[227] R. Renthal. Helix insertion into bilayers and the evolution of membrane proteins. *Cell. Mol. Life. Sci.*, 67:1077–1088, 2010.

[228] M.B. Ulmschneider and M.S.P. Sansom. Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta*, 1512:1–14, 2001.

[229] M. Gassmann, B. Grenacher, B. Rohde, and J. Vogel. Quantifying western blots: Pitfalls and densitometry. *Electrophoresis*, 30:1845–1855, 2009.

[230] F.X. Zhou, M.J. Cocco, W.P. Russ, A.T. Brunger, and D.M. Engelman. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.*, 7:154–160, 2000.

[231] J.U. Bowie. Membrane protein folding: how important are hydrogen bonds? *Curr. Opin. Struct. Biol.*, 21:42–49, 2011.

[232] R. Taylor, O. Kennard, and W. Versichel. Geometry of the imino-carbonyl (nh-o=c) hydrogen bond. 1. lone-pair directionality. *J. Am. Chem. Soc.*, 105:5761–5766, 1984.

[233] R. Taylor, O. Kennard, and W. Versichel. Geometry of the imino-carbonyl (nh-o=c) hydrogen bond. 3. hydrogen bond distances and angles. *Acta Crystallogr., Sect. B*, 40:280–288, 1984.

[234] T. M. Gray and B. W. Matthews. Interhelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.*, 175:75–81, 1984.

[235] J.A. Ballesteros, X. Deupi, M. Olivella, E.E. Haaksma, and L. Pardo. Serine and threonine residues bend alpha-helices in the chi(1)=g(-) conformation. *Biophys J.*, 79:2754–2760, 2000.

[236] F.X. Zhou, H.J. Merianos, A.T. Brunger, and D.M. Engelman. Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl. Acad. Sci. U.S.A.*, 98(5):2250–2255, 2006.

[237] J. Nordholm, D.V. da Silva, J. Damjanovic, D. Dou, and R. Daniels. Polar residues and their positional context dictate the transmembrane domain interactions of influenza a neuraminidases. *Biochemistry*, 48(40):9437–9447, 2009.

[238] K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman. A transmembrane helix dimer: structure and implications. *Science*, 276:131–133, 1997.

[239] S. O. Smith, D. Song, S. Shekar, M. Groesbeek, M. Ziliox, and S. Aimoto. Structure of the transmembrane dimer interface of glycophorin a in membrane bilayers. *Biochemistry*, 40:6553–6558, 2001.

[240] A. Senes, I. Ubarretxena-Belandia, and D. M. Engelman. The c_α-h·o hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. U.S.A.*, 98:9056–9061, 2001.

[241] P. Cosson and J. S. Bonifacino. Role of transmembrane domain interactions in the assembly of class ii mhc molecules. *Science*, 258:659–662, 1992.

[242] K. R. MacKensie and D. M. Engelman. Structure-based prediction of the stability of transmembrane helix-helix interactions: the sequence dependence of glycophorin a dimerization. *Proc. Natl. Acad. Sci. U.S.A.*, 95:3583–3590, 1998.

[243] E Sackmann. Biological membranes architecture and function. *J. Biomol. Struct. Dyn.*, pages 1–65, 1995.

[244] J.A. Killian and B. Dekruijff. Thermodynamic, motional, and structural aspects of the gramicidin-induced hexagonal hii phase formation in phophatidylethanolamine. *Biochemistry*, 24:7881–7890, 1985.

[245] E. Wallin and G von Heijne. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, 7:1029–1038, 1998.

[246] E. Strandberg and J. A. Killian. Snorkeling of lysine side chains in transmembrane helices: how easy can it get? *FEBS Lett.*, 544:69–73, 2003.

[247] S.K. Kandasamy and R.G. Larson. Molecular dynamics simulations of model *Trans*-membrane peptides in lipid bilayers: A systematic investigation of hydrophobic mismatch. *Biophys. J.*, 90:2326–2343, 2006.

[248] D. E. Elmore and D. A. Dougherty. Investigating lipid composition effects on the mechanosensitive channel of large conductance (mscl) using molecular dynamics simulations. *Biophys J.*, 85:685–740, 2003.

[249] M. Venturoli, B. Smit, and M.M. Sperotto. Simulation studies of protein-induced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins. *Biophys. J.*, 88:1778–1798, 2005.

[250] M. Yiannourakou, L. Marsella, F. de Meyer, and B. Smit. Towards an understanding of membrane-mediated protein-protein interactions. *Faraday Discuss.*, 144:359–367, 2010.

[251] I.H. Al-Lehyani, J.M.A. Grime, M. Bano, K. McKelvey, and M.P. Allen. Coarse-grained simulation of transmembrane peptides in the gel phase. *J. Comput. Phys.*, 238:97–105, 2013.

[252] M. Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, 1996.

[253] N. Kuĉerka, M. P. Nieh, and J. Katsaras. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim. Biophys. Acta*, 1808:2761–2771, 2011.

[254] S.O. Nielsen, B. Ensing, V. Oritz, P.B. Moore, and M.L. Klein. Lipid bilayer perturbations around a transmembrane nanotube: A coarse grain molecular dynamics study. *Biophys. J.*, 88:3822–2828, 2005.

[255] L. Li, I. Vorobyov, A.D. Mackerell, and T.W. Allen. Is arginine charged in a membrane? *Biophys J.*, 94(2):L11–L13, 2008.

[256] L-P. Liu and C. M. Deber. Uncoupling hydrophobicity and helicity in transmembrane segments. *J. Biol. Chem.*, 273(37):23645–23648, 1998.

[257] G.E. Tusnády, Z. Dosztányi, and I. Simon. Transmembrane proteins in the protein data bank: identification and classification. *Bioinformatics*, 20(17):2964–2972, 2004.

[258] G.E. Tusnády, L. Kalmár, and Simon I. Topdb: topology data bank of transmembrane proteins. *Nucleic Acids Res.*, 36:D234–D239, 2008.

[259] S. A. Fernando, P. Selvarani, Soma Das, Ch. Kiran Kumar, Sukanta Mondal, S. Ramakumar, and K. Sekar. Thgs: a web?based database of transmembrane helices in genome sequences. *Nucleic Acids Res.*, 32(suppl 1):D125–D128, 2004.

[260] A. Lo, C.W. Cheng, Y.Y. Chiu, T.Y. Sung, and W.L. Hsu. Tmpad: an integrated structural database for helix-packing folds in transmembrane proteins. *Nucleic Acids Res.*, 39:D347–D355, 2011.

[261] O.D. Testa, E. Moutevelis, and D.N. Woolfson. Cc+: a relational database of coiled-coil structures. *Nucleic Acids Res.*, 37:D315–D322, 2009.

[262] J. Walshaw and D.N. Woolfson. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.*, 307:1427–1450, 2001.

[263] C. Chothia, M. Levitt, and D. Richardson. Helix to helix packing in proteins. *J. Mol. Biol.*, 145:215–250, 1981.

[264] J. Lee and W. Im. Implementation and application of helix-helix distance and crossing angle restraint potentials. *J. Comput. Chem*, 28:669–680, 2007.

Appendix A

Helix-helix crossing angle and tilt angle

The calculation of dimer crossing angles and individual helix tilt angles were generated using a MATLAB implementation of the following mathematics. They are based on the definitions by Chothia *et al.,* (263), and further refined through implementation in PMF calculations (264). To begin with the minimum distance between helices was determined by first populating an inertia tensor to yield the principal axis for both helices (section A.1). An equation of two interesting straight lines was generating by attaining two 'shifting' arbitrary points along each principal axis between a predefined beginning and ending point of geometry. The two points are minimised by taking the second differential of their difference with respect to a 'shifting' constant. The helix-helix crossing angle (section A.2) and tilt angle of individual helices (section A.3) can be calculated independently from the calculations performed to define the minimum *inter*helical distance.

A.1 Minimum distance between helices

The principal axis $\boldsymbol{a}^{(k)}$ of helix k is represented by the eigenvector with the smallest eigenvalue from the inertia tensor $\boldsymbol{M}^{(k)}$. The matrix elements, M_{ij} are defined by:

$$M_{ij}^{(k)} - \sum_{\alpha}^{n} (P_{i,\alpha}^{(k)} - \langle P_{i}^{(k)} \rangle)(P_{j,\alpha}^{(k)} - \langle P_{j}^{(k)} \rangle), \tag{A.1}$$

where n is the number of alpha carbon atoms along the back bone of helix k. The elements of vector $\boldsymbol{P}_{\alpha}^{(k)}$ is defined as:

$$P_{i,\alpha} \in \boldsymbol{P}_{\alpha}^{(k)} | \boldsymbol{P}_{\alpha}^{(k)} = \boldsymbol{r}_{\alpha}^{(k)} + \boldsymbol{r}_{\alpha+2}^{(k)} - 2\boldsymbol{r}_{\alpha-1}^{(k)}, \tag{A.2}$$

where $i = x,\ y,\ z$, and $\boldsymbol{r}_{\alpha}^{(k)}$ is a set of cartesian coordinates of the α-carbon, α, on helix k. The average position $\langle P^{(k)} \rangle$ is defined as:

$$\langle \boldsymbol{P}^{(k)} \rangle = \frac{1}{n} \sum_{\alpha} \boldsymbol{P}_{\alpha}^{(k)}. \tag{A.3}$$

To define the helix beginning ($\boldsymbol{b}^{(k)}$) and ending ($\boldsymbol{e}^{(k)}$):

$$\boldsymbol{b}^{k} = \boldsymbol{r}_{c}^{(k)} + [\boldsymbol{a}^{(k)}(\boldsymbol{r}_{l}^{(k)} - \boldsymbol{r}_{c}^{(k)})]\boldsymbol{a}^{k} \tag{A.4}$$

and

$$\boldsymbol{e}^{k} = \boldsymbol{r}_{c}^{(k)} + [\boldsymbol{a}^{(k)}(\boldsymbol{r}_{n}^{(k)} - \boldsymbol{r}_{c}^{(k)})]\boldsymbol{a}^{k}, \tag{A.5}$$

where $\boldsymbol{r}_{c}^{(k)}$ is the centre of geometry of helix k. An arbitrary point $\boldsymbol{t}^{(k)}$ along helix k can be defined as:

$$\boldsymbol{t}^{(k)} = \boldsymbol{b}^{(k)} + S^{(k)}(\boldsymbol{e}^{(k)} - \boldsymbol{b}^{(k)}), \tag{A.6}$$

where $S^{(k)}$ is a constant $S \in (0, 1)$. The shortest distance between two helices k and j is defined by minimising the square of the distance between $\boldsymbol{t}^{(k)}$ and $\boldsymbol{t}^{(j)}$:

$$D = |\boldsymbol{t}^{(k)} - \boldsymbol{t}^{(j)}|. \tag{A.7}$$

The minimum distance can be obtained by the second differential of D with respect to $S^{(k)}$ and $S^{(j)}$:

$$\frac{\partial D^2}{\partial S^{(k)}} = (\boldsymbol{b}^{(k)} - \boldsymbol{b}^{(j)})(\boldsymbol{e}^{(k)} - \boldsymbol{b}^{(k)}) - (\boldsymbol{e}^{(j)} - \boldsymbol{b}^{(j)})(\boldsymbol{e}^{(k)} - \boldsymbol{b}^{(k)})S^{(j)} + |\boldsymbol{e}^{(k)} - \boldsymbol{b}^{(k)}|^2 S^{(k)} = 0, \tag{A.8}$$

and

$$\frac{\partial D^2}{\partial S^{(j)}} = (\boldsymbol{b}^{(k)} - \boldsymbol{b}^{(j)})(\boldsymbol{e}^{(j)} - \boldsymbol{b}^{(j)}) - (\boldsymbol{e}^{(k)} - \boldsymbol{b}^{(k)})(\boldsymbol{e}^{(j)} - \boldsymbol{b}^{(j)})S^{(k)} + |\boldsymbol{e}^{(j)} - \boldsymbol{b}^{(j)}|^2 S^{(j)} = 0. \tag{A.9}$$

A.2 Helix-helix crossing angle

The crossing angle, θ, of two helices k and j is defined as the dihedral angle of two planes (Figure A.1):

$$\theta_{cr} = \cos^{-1}\Big(\frac{(\boldsymbol{l} \times \boldsymbol{h}) \cdot (\boldsymbol{h} \times \boldsymbol{m})}{|\boldsymbol{l} \times \boldsymbol{h}||\boldsymbol{h} \times \boldsymbol{m}|}\Big), \tag{A.10}$$

where $\boldsymbol{h} = (\boldsymbol{t}^{(j)} - \boldsymbol{t}^{(k)})/|\boldsymbol{t}^{(j)} - \boldsymbol{t}^{(k)}|$, $\boldsymbol{l} = (\boldsymbol{t}^{(k)} - \boldsymbol{b}^{(k)})/|\boldsymbol{t}^{(k)} - \boldsymbol{b}^{(k)}| = \boldsymbol{a}^{(k)}$, and $\boldsymbol{m} = -\boldsymbol{a}^{(j)}$.

A.3 Helix tilt angle

The tilt angle is defined as the angle between the principal axis of the helix and the unit vector of the bilayer normal (see Figure A.1). The curvature in a bilayer 'slab' is considered flat over time, therefore the bilayer normal, typically z, is defined as the constant vector $\boldsymbol{u} = (0, 0, 1)$. The angle is defined as:

$$\theta_{tl} = \cos(\boldsymbol{u} \cdot \boldsymbol{a}^{(k)}). \tag{A.11}$$

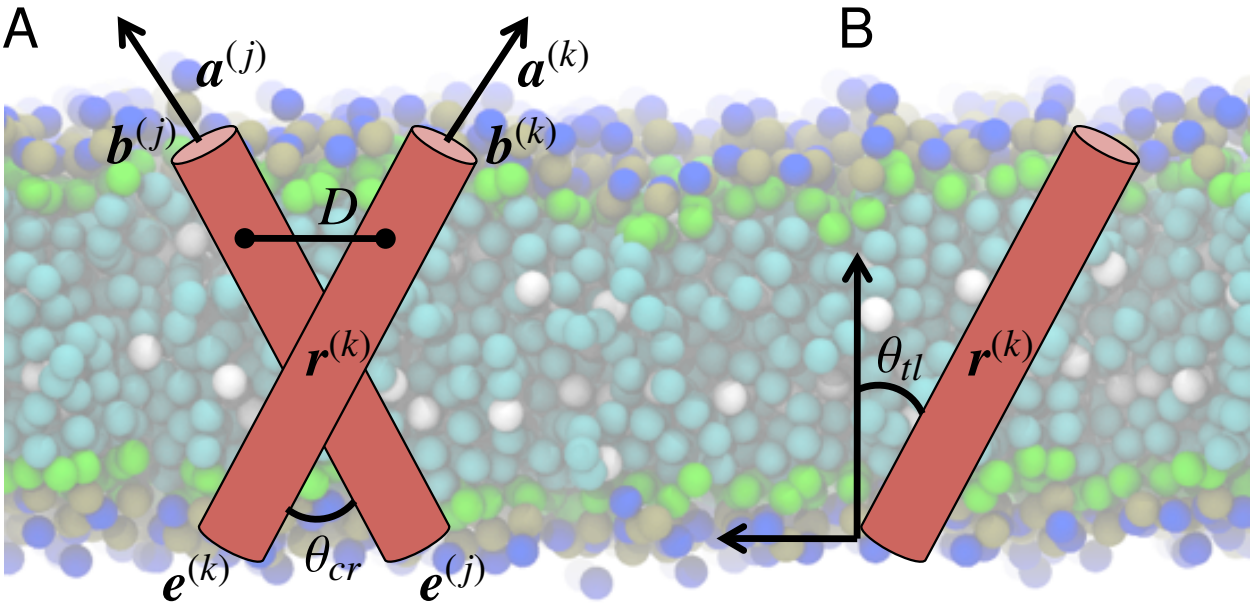

Figure A.1: Helix-helix crossing angle (A) and individual helix tilt angle (B). The helix-helix (or dimer) crossing angle is defined as the angle between the principal axis of each helix. The tilt angle can be calculated by taking the angle between the principal axis of the helix and the bilayer normal. The variables are defined in the text.